

Policy and Theory of International Economics

v. 1.0

This is the book [Policy and Theory of International Economics](#) (v. 1.0).

This book is licensed under a [Creative Commons by-nc-sa 3.0](#) (<http://creativecommons.org/licenses/by-nc-sa/3.0/>) license. See the license for more details, but that basically means you can share this book as long as you credit the author (but see below), don't make money from it, and do make it available to everyone else under the same terms.

This book was accessible as of December 29, 2012, and it was downloaded then by [Andy Schmitz](#) (<http://lardbucket.org>) in an effort to preserve the availability of this book.

Normally, the author and publisher would be credited here. However, the publisher has asked for the customary Creative Commons attribution to the original publisher, authors, title, and book URI to be removed. Additionally, per the publisher's request, their name has been removed in some passages. More information is available on this project's [attribution page](#) (http://2012books.lardbucket.org/attribution.html?utm_source=header).

For more information on the source of this book, or why it is available for free, please see [the project's home page](#) (<http://2012books.lardbucket.org/>). You can browse or download additional books there.

Table of Contents

About the Author	1
Acknowledgments	2
Preface.....	3
Chapter 1: Introductory Trade Issues: History, Institutions, and Legal Framework	4
The International Economy and International Economics	5
Understanding Tariffs	11
Recent Trade Controversies.....	17
The Great Depression, Smoot-Hawley, and the Reciprocal Trade Agreements Act (RTAA)	26
The General Agreement on Tariffs and Trade (GATT)	29
The Uruguay Round.....	39
The World Trade Organization	46
Appendix A: Selected U.S. Tariffs—2009	51
Appendix B: Bound versus Applied Tariffs.....	59
Chapter 2: The Ricardian Theory of Comparative Advantage	62
The Reasons for Trade	63
The Theory of Comparative Advantage: Overview.....	66
Ricardian Model Assumptions	77
The Ricardian Model Production Possibility Frontier	84
Definitions: Absolute and Comparative Advantage.....	87
A Ricardian Numerical Example	95
Relationship between Prices and Wages	103
Deriving the Autarky Terms of Trade.....	105
The Motivation for International Trade and Specialization	108
Welfare Effects of Free Trade: Real Wage Effects	114
The Welfare Effects of Free Trade: Aggregate Effects	122
Appendix: Robert Torrens on Comparative Advantage	125

Chapter 3: The Pure Exchange Model of Trade	127
A Simple Pure Exchange Economy	128
Determinants of the Terms of Trade	131
Example of a Trade Pattern	136
Three Traders and Redistribution with Trade	139
Three Traders with International Trade	144
The Nondiscrimination Argument for Free Trade	146
Chapter 4: Factor Mobility and Income Redistribution.....	150
Factor Mobility Overview	151
Domestic Factor Mobility	154
Time and Factor Mobility	157
Immobile Factor Model Overview and Assumptions.....	160
The Production Possibility Frontier in the Immobile Factor Model	165
Autarky Equilibrium in the Immobile Factor Model.....	168
Depicting a Free Trade Equilibrium in the Immobile Factor Model.....	172
Effect of Trade on Real Wages	175
Intuition of Real Wage Effects	179
Interpreting the Welfare Effects	181
Aggregate Welfare Effects of Free Trade in the Immobile Factor Model	183
Chapter 5: The Heckscher-Ohlin (Factor Proportions) Model	186
Chapter Overview.....	187
Heckscher-Ohlin Model Assumptions.....	196
The Production Possibility Frontier (Fixed Proportions)	205
The Rybczynski Theorem	209
The Magnification Effect for Quantities	212
The Stolper-Samuelson Theorem.....	219
The Magnification Effect for Prices.....	223
The Production Possibility Frontier (Variable Proportions)	230
The Heckscher-Ohlin Theorem.....	234
Depicting a Free Trade Equilibrium in the Heckscher-Ohlin Model	241
National Welfare Effects of Free Trade in the Heckscher-Ohlin Model	244
The Distributive Effects of Free Trade in the Heckscher-Ohlin Model	247
The Compensation Principle.....	253
Factor-Price Equalization.....	258
The Specific Factor Model: Overview.....	262
The Specific Factor Model	268
Dynamic Income Redistribution and Trade	283

Chapter 6: Economies of Scale and International Trade	298
Chapter Overview.....	299
Economies of Scale and Returns to Scale	301
Gains from Trade with Economies of Scale: A Simple Explanation	305
Monopolistic Competition	312
Model Assumptions: Monopolistic Competition.....	314
The Effects of Trade in a Monopolistically Competitive Industry	317
The Costs and Benefits of Free Trade under Monopolistic Competition	323
Chapter 7: Trade Policy Effects with Perfectly Competitive Markets	328
Basic Assumptions of the Partial Equilibrium Model.....	329
Depicting a Free Trade Equilibrium: Large and Small Country Cases	332
The Welfare Effects of Trade Policies: Partial Equilibrium	341
Import Tariffs: Large Country Price Effects	350
Import Tariffs: Large Country Welfare Effects	358
The Optimal Tariff.....	367
Import Tariffs: Small Country Price Effects	372
Import Tariffs: Small Country Welfare Effects	375
Retaliation and Trade Wars	381
Import Quotas: Large Country Price Effects.....	391
Administration of an Import Quota	395
Import Quota: Large Country Welfare Effects	398
Import Quota: Small Country Price Effects	406
Import Quota: Small Country Welfare Effects.....	409
The Choice between Import Tariffs and Quotas	415
Export Subsidies: Large Country Price Effects	423
Export Subsidies: Large Country Welfare Effects	426
Countervailing Duties	433
Voluntary Export Restraints (VERs): Large Country Price Effects	443
Administration of a Voluntary Export Restraint.....	446
Voluntary Export Restraints: Large Country Welfare Effects	449
Export Taxes: Large Country Price Effects	456
Export Taxes: Large Country Welfare Effects	459

Chapter 8: Domestic Policies and International Trade	465
Chapter Overview.....	466
Domestic Production Subsidies	473
Production Subsidies as a Reason for Trade	475
Production Subsidy Effects in a Small Importing Country.....	478
Domestic Consumption Taxes.....	483
Consumption Taxes as a Reason for Trade	485
Consumption Tax Effects in a Small Importing Country	488
Equivalence of an Import Tariff with a Domestic (Consumption Tax plus Production Subsidy)	493
Chapter 9: Trade Policies with Market Imperfections and Distortions	499
Chapter Overview.....	500
Imperfections and Distortions Defined	504
The Theory of the Second Best	511
Unemployment and Trade Policy.....	517
The Infant Industry Argument and Dynamic Comparative Advantage	529
The Case of a Foreign Monopoly	541
Monopoly and Monopsony Power and Trade	549
Public Goods and National Security	556
Trade and the Environment.....	563
Economic Integration: Free Trade Areas, Trade Creation, and Trade Diversion	578
Chapter 10: Political Economy and International Trade	592
Chapter Overview.....	593
Some Features of a Democratic Society.....	596
The Economic Effects of Protection: An Example.....	599
The Consumers' Lobbying Decision	602
The Producers' Lobbying Decision	605
The Government's Decision	608
The Lobbying Problem in a Democracy	611
Chapter 11: Evaluating the Controversy between Free Trade and Protectionism.....	614
Introduction	615
Economic Efficiency Effects of Free Trade	618
Free Trade and the Distribution of Income	621
The Case for Selected Protection.....	626
The Economic Case against Selected Protection	631
Free Trade as the “Pragmatically Optimal” Policy Choice	640

Chapter 12: Introductory Finance Issues: Current Patterns, Past History, and International Institutions	642
GDP, Unemployment, Inflation, and Government Budget Balances	643
Exchange Rate Regimes, Trade Balances, and Investment Positions	653
Business Cycles: Economic Ups and Downs	662
International Macroeconomic Institutions: The IMF and the World Bank	672
Chapter 13: National Income and the Balance of Payments Accounts	679
National Income and Product Accounts	680
National Income or Product Identity	686
U.S. National Income Statistics (2007–2008)	692
Balance of Payments Accounts: Definitions	696
Recording Transactions on the Balance of Payments	703
U.S. Balance of Payments Statistics (2008)	713
The Twin-Deficit Identity	723
International Investment Position	740
Chapter 14: The Whole Truth about Trade Imbalances	745
Overview of Trade Imbalances	746
Trade Imbalances and Jobs	748
The National Welfare Effects of Trade Imbalances	753
Some Further Complications	772
How to Evaluate Trade Imbalances	775
Chapter 15: Foreign Exchange Markets and Rates of Return	793
The Forex: Participants and Objectives	794
Exchange Rate: Definitions	798
Calculating Rate of Returns on International Investments	805
Interpretation of the Rate of Return Formula	809
Applying the Rate of Return Formulas	815
Chapter 16: Interest Rate Parity	821
Overview of Interest Rate Parity	822
Comparative Statics in the IRP Theory	826
Forex Equilibrium with the Rate of Return Diagram	833
Exchange Rate Equilibrium Stories with the RoR Diagram	836
Exchange Rate Effects of Changes in U.S. Interest Rates Using the RoR Diagram	840
Exchange Rate Effects of Changes in Foreign Interest Rates Using the RoR Diagram	843
Exchange Rate Effects of Changes in the Expected Exchange Rate Using the RoR Diagram	847

Chapter 17: Purchasing Power Parity	851
Overview of Purchasing Power Parity (PPP)	852
The Consumer Price Index (CPI) and PPP	857
PPP as a Theory of Exchange Rate Determination	861
Problems and Extensions of PPP	867
PPP in the Long Run.....	871
Overvaluation and Undervaluation	876
PPP and Cross-Country Comparisons	882
Chapter 18: Interest Rate Determination	886
Overview of Interest Rate Determination	887
Some Preliminaries.....	890
What Is Money?	893
Money Supply Measures.....	896
Controlling the Money Supply.....	900
Money Demand.....	906
Money Functions and Equilibrium	910
Money Market Equilibrium Stories.....	914
Effects of a Money Supply Increase.....	918
Effect of a Price Level Increase (Inflation) on Interest Rates	921
Effect of a Real GDP Increase (Economic Growth) on Interest Rates.....	924
Integrating the Money Market and the Foreign Exchange Markets	927
Comparative Statics in the Combined Money-Forex Model	932
Money Supply and Long-Run Prices	937
Chapter 19: National Output Determination.....	945
Overview of National Output Determination.....	946
Aggregate Demand for Goods and Services.....	950
Consumption Demand	952
Investment Demand.....	956
Government Demand.....	958
Export and Import Demand	960
The Aggregate Demand Function	964
The Keynesian Cross Diagram	966
Goods and Services Market Equilibrium Stories.....	969
Effect of an Increase in Government Demand on Real GNP	974
Effect of an Increase in the U.S. Dollar Value on Real GNP	977
The J-Curve Effect	980

Chapter 20: The AA-DD Model.....	986
Overview of the AA-DD Model	987
Derivation of the DD Curve	990
Shifting the DD Curve	995
Derivation of the AA Curve	998
Shifting the AA Curve	1003
Superequilibrium: Combining DD and AA	1007
Adjustment to the Superequilibrium.....	1011
AA-DD and the Current Account Balance	1017
Chapter 21: Policy Effects with Floating Exchange Rates.....	1023
Overview of Policy with Floating Exchange Rates.....	1024
Monetary Policy with Floating Exchange Rates	1028
Fiscal Policy with Floating Exchange Rates.....	1034
Expansionary Monetary Policy with Floating Exchange Rates in the Long Run.....	1041
Foreign Exchange Interventions with Floating Exchange Rates	1047
Chapter 22: Fixed Exchange Rates.....	1055
Overview of Fixed Exchange Rates	1056
Fixed Exchange Rate Systems	1059
Interest Rate Parity with Fixed Exchange Rates	1070
Central Bank Intervention with Fixed Exchange Rates	1073
Balance of Payments Deficits and Surpluses	1077
Black Markets	1080
Chapter 23: Policy Effects with Fixed Exchange Rates	1083
Overview of Policy with Fixed Exchange Rates	1084
Monetary Policy with Fixed Exchange Rates	1088
Fiscal Policy with Fixed Exchange Rates	1093
Exchange Rate Policy with Fixed Exchange Rates	1099
Reserve Country Monetary Policy under Fixed Exchange Rates	1106
Currency Crises and Capital Flight.....	1111
Case Study: The Breakup of the Bretton Woods System, 1973	1117
Chapter 24: Fixed versus Floating Exchange Rates	1128
Overview of Fixed versus Floating Exchange Rates	1129
Exchange Rate Volatility and Risk	1131
Inflationary Consequences of Exchange Rate Systems	1137
Monetary Autonomy and Exchange Rate Systems	1141
Which Is Better: Fixed or Floating Exchange Rates?	1145

About the Author

Steve Suranovic

Steve Suranovic is an associate professor of economics and international affairs at the George Washington University (GW) in Washington, DC. He has a PhD in economics from Cornell University and a BS in mathematics from the University of Illinois at Urbana-Champaign. He has been teaching international trade and finance for more than twenty years at GW and as an adjunct for Cornell University's Washington, DC, program. In fall 2002, he taught at Sichuan University in Chengdu, China, as a visiting Fulbright lecturer. He has taught a GW class at Fudan University in Shanghai during the summers of 2009 and 2010. He has also spoken to business, government, and academic audiences in Japan, Malaysia, the Philippines, China, and Mongolia as part of the U.S. State Department speaker's programs.



His research focuses on two areas: international trade policy and behavioral economics. With respect to behavior, he examines why people choose to do things that many observers view as irrational. Examples include addiction to cigarettes, cyclical dieting, and anorexia. His research shows that dangerous behaviors can be explained as the outcome of a reasoned and rational optimization exercise. With respect to trade policy, his research seeks to reveal the strengths and weaknesses of arguments supporting various policy options. The goal is to answer the question, what trade policies should a country implement? More generally, he applies the economic analytical method to identify the policies that can attract the most widespread support.

His book *A Moderate Compromise: Economic Policy Choice in an Era of Globalization* will be released by Palgrave Macmillan in fall 2010. In it he offers a critique of current methods to evaluate and choose policies and suggests a simple, principled, and moderate alternative.

He also blogs occasionally at <http://stevesuranovic.blogspot.com>.

Acknowledgments

I am most indebted to my students at the George Washington University, Cornell University, and Sichuan University and the visitors at the International Economics Study Center Web site. Students during the past twenty-plus years and Web site visitors for the past ten-plus years have been the primary audience for these writings. Nothing has been more encouraging than hearing a student express how much more intelligible are economics news stories in the *Wall Street Journal* or *Financial Times* after taking one of my courses or receiving an e-mail about how helpful the freely available online notes have been. I thank all those students and readers for their encouraging remarks.

I am also indebted to my teachers, going back to the primary school teachers at Assumption BVM in Chicago (especially Sister Marie), high school teachers at Lincoln-Way in New Lenox, Illinois (especially Bill Colgan), professors at the University of Illinois at Urbana-Champaign, and my economics professors at Cornell University (especially Henry Wan, George Staller, Jan Svejnar, David Easley, Mukul Majumdar, Tapan Mitra, Earl Grinols, Gary Fields, and Robert Frank).

For my teaching style, I am grateful to my teachers via textbooks, including William Baumol, Alan Blinder, Hal Varian, Paul Krugman, Maurice Obstfeld, and especially Eugene Silberberg, whose graduate-level book *The Structure of Economics*, offering detailed and logical explanations of economic models, was most illuminating and inspiring.

I am also grateful to my colleagues at GW, all of whom have contributed in numerous ways via countless conversations about economic issues through the years. Particular students who have contributed to the Flat World edition include Runping Xu and Jiyoung Lee. Finally, I am thankful to the reviewers and production staff from Unnamed Publisher.

On a personal note, I remain continually grateful for the loving support of my family; my children, Ben and Katelyn; and M. Victoria Farrales.

Preface

Traditionally, intermediate-level international economics texts seem to fall into one of two categories. Some are written for students who may one day continue on in an economics PhD program. These texts develop advanced general equilibrium models and use sophisticated mathematics. However, these texts are also very difficult for the average, non-PhD-bound student to understand. Other intermediate texts are written for noneconomics majors who may take only a few economics courses in their program. These texts present descriptive information about the world and only the bare basics about how economic models are used to describe that world.

This text strives to reach a median between these two approaches. First, I believe that students need to learn the theory and models to understand how economists understand the world. I also think these ideas are accessible to most students if they are explained thoroughly. This text presents numerous models in some detail, not by employing advanced mathematics, but rather by walking students through a detailed description of how a model's assumptions influence its conclusions.

Second, and perhaps more important, students must learn how the models connect with the real world. I believe that theory is done primarily to guide policy. We do positive economics to help answer the normative questions; for example, what should a country do about its trade policy or its exchange rate policy? The results from models give us insights that help us answer these questions. Thus this text strives to explain why each model is interesting by connecting its results to some aspect of a current policy issue. A prime example is found in [Chapter 11 "Evaluating the Controversy between Free Trade and Protectionism"](#) of this book, which addresses the age-old question of whether countries should choose free trade or some type of selected protection. The chapter demonstrates how the results of the various models presented throughout the text contribute to our understanding of this long-standing debate.

Chapter 1

Introductory Trade Issues: History, Institutions, and Legal Framework

Economics is a social science whose purpose is to understand the workings of the real-world economy. An economy is something that no one person can observe in its entirety. We are all a part of the economy, we all buy and sell things daily, but we cannot observe all parts and aspects of an economy at any one time.

For this reason, economists build mathematical models, or theories, meant to describe different aspects of the real world. For some students, economics seems to be all about these models and theories, these abstract equations and diagrams. However, in actuality, economics is about the real world, the world we all live in.

For this reason, it is important in any economics course to describe the conditions in the real world before diving into the theory intended to explain them. In this case, in a textbook about international trade, it is very useful for a student to know some of the policy issues, the controversies, the discussions, and the history of international trade.

This first chapter provides an overview of the real world with respect to international trade. It explains not only where we are now but also where we have been and why things changed along the way. It describes current trade laws and institutions and explains why they have been implemented.

With this overview about international trade in the real world in mind, a student can better understand why the theories and models in the later chapters are being developed. This chapter lays the groundwork for everything else that follows.

1.1 The International Economy and International Economics

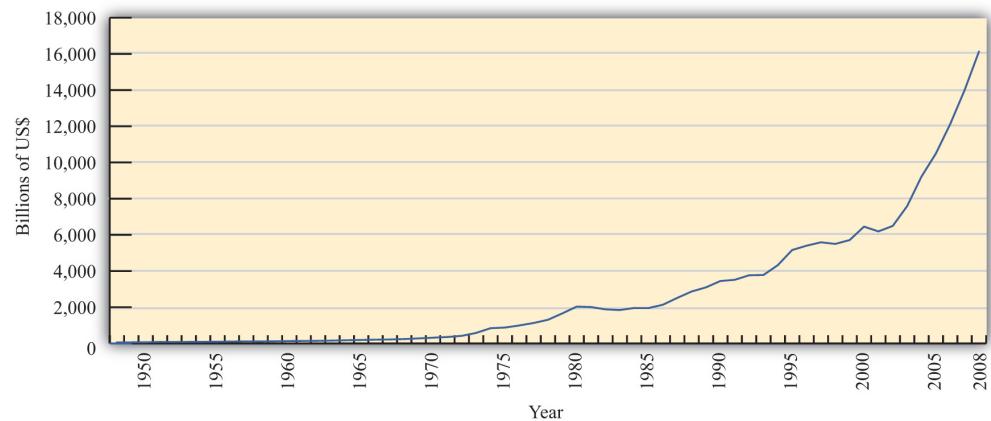
LEARNING OBJECTIVES

1. Learn past trends in international trade and foreign investment.
2. Learn the distinction between international trade and international finance.

International economics is growing in importance as a field of study because of the rapid integration of international economic markets. Increasingly, businesses, consumers, and governments realize that their lives are affected not only by what goes on in their own town, state, or country but also by what is happening around the world. Consumers can walk into their local shops today and buy goods and services from all over the world. Local businesses must compete with these foreign products. However, many of these same businesses also have new opportunities to expand their markets by selling to a multitude of consumers in other countries. The advance of telecommunications is also rapidly reducing the cost of providing services internationally, while the Internet will assuredly change the nature of many products and services as it expands markets even further.

One simple way to see the rising importance of international economics is to look at the growth of exports in the world during the past fifty or more years. [Figure 1.1 "World Exports, 1948–2008 \(in Billions of U.S. Dollars\)"](#) shows the overall annual exports measured in billions of U.S. dollars from 1948 to 2008. Recognizing that one country's exports are another country's imports, one can see the exponential growth in outflows and inflows during the past fifty years.

Figure 1.1 World Exports, 1948–2008 (in Billions of U.S. Dollars)



Source: World Trade Organization, International trade and tariff data, http://www.wto.org/english/res_e/statistics_e/statistics_e.htm.

However, rapid growth in the value of exports does not necessarily indicate that trade is becoming more important. A better method is to look at the share of traded goods in relation to the size of the world economy. Figure 1.2 "World Exports, 1970–2008 (Percentage of World GDP)" shows world exports as a percentage of the world gross domestic product (GDP) for the years 1970 to 2008. It shows a steady increase in trade as a share of the size of the world economy. World exports grew from just over 10 percent of the GDP in 1970 to over 30 percent by 2008. Thus trade is not only rising rapidly in absolute terms; it is becoming relatively more important too.

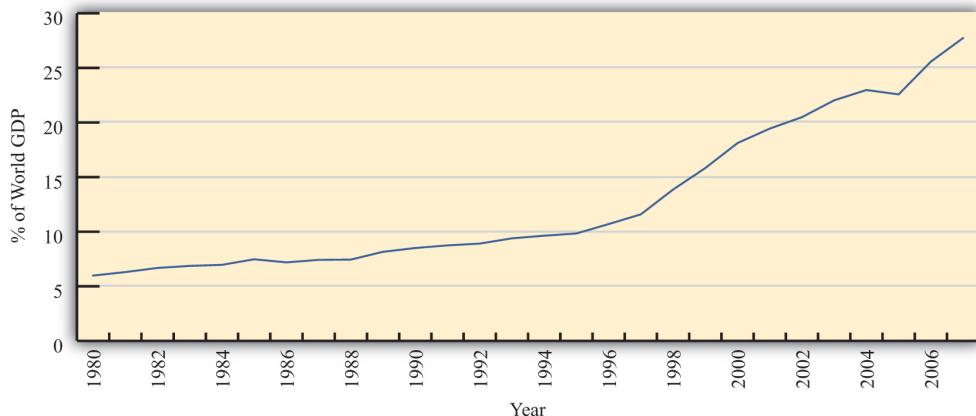
Figure 1.2 World Exports, 1970–2008 (Percentage of World GDP)



Source: IMF World Economic Outlook Database, <http://www.imf.org/external/pubs/ft/weo/2009/02/weodata/index.aspx>.

One other indicator of world interconnectedness can be seen in changes in the amount of foreign direct investment (FDI). FDI is foreign ownership of productive activities and thus is another way in which foreign economic influence can affect a country. Figure 1.3 "World Inward FDI Stocks, 1980–2007 (Percentage of World GDP)" shows the stock, or the sum total value, of FDI around the world taken as a percentage of the world GDP between 1980 and 2007. It gives an indication of the importance of foreign ownership and influence around the world. As can be seen, the share of FDI has grown dramatically from around 5 percent of the world GDP in 1980 to over 25 percent of the GDP just twenty-five years later.

Figure 1.3 *World Inward FDI Stocks, 1980–2007 (Percentage of World GDP)*



Source: IMF World Economic Outlook Database, <http://www.imf.org/external/pubs/ft/weo/2009/02/weodata/index.aspx>; UNCTAD, FDI Statistics: Division on Investment and Enterprise, <http://www.unctad.org/Templates/Page.asp?intItemID=4979&lang=1>.

1. An international agreement among countries, established in 1948, promoting trade liberalization through the reduction of tariff rates and other barriers to trade until its conversion to the WTO in 1995.
2. The eighth and last round of GATT trade liberalization negotiations that substantially expanded the number and scope of trade liberalization agreements and established the WTO.
3. An international agency whose purpose is to monitor and enforce the Uruguay Round trade liberalization agreements and to promote continuing liberalizing initiatives with continuing rounds of negotiation.

The growth of international trade and investment has been stimulated partly by the steady decline of trade barriers since the Great Depression of the 1930s. In the post-World War II era, the **General Agreement on Tariffs and Trade**¹, or GATT, prompted regular negotiations among a growing body of members to reciprocally reduce tariffs (import taxes) on imported goods. During each of these regular negotiations (eight of these rounds were completed between 1948 and 1994), countries promised to reduce their tariffs on imports in exchange for concessions—that means tariffs reductions—by other GATT members. When the **Uruguay Round**², the most recently completed round, was finalized in 1994, the member countries succeeded in extending the agreement to include liberalization promises in a much larger sphere of influence. Now countries not only would lower tariffs on goods trade but also would begin to liberalize the agriculture and services markets. They would eliminate the many quota systems—like the multifiber agreement in clothing—that had sprouted up in previous decades. And they would agree to adhere to certain minimum standards to protect intellectual property rights such as patents, trademarks, and copyrights. The **World Trade Organization (WTO)**³ was created to manage this system of new agreements, to provide a forum for regular discussion of trade matters, and to implement a well-defined process for settling trade disputes that might arise among countries.

As of 2009, 153 countries were members of the WTO “trade liberalization club,” and many more countries were still negotiating entry. As the club grows to include more members—and if the latest round of trade liberalization talks, called the Doha Round, concludes with an agreement—world markets will become increasingly

open to trade and investment. Note that the Doha Round of discussions was begun in 2001 and remains uncompleted as of 2009.

Another international push for trade liberalization has come in the form of regional free trade agreements. Over two hundred regional trade agreements around the world have been notified, or announced, to the WTO. Many countries have negotiated these agreements with neighboring countries or major trading partners to promote even faster trade liberalization. In part, these have arisen because of the slow, plodding pace of liberalization under the GATT/WTO. In part, the regional trade agreements have occurred because countries have wished to promote interdependence and connectedness with important economic or strategic trade partners. In any case, the phenomenon serves to open international markets even further than achieved in the WTO.

These changes in economic patterns and the trend toward ever-increasing openness are an important aspect of the more exhaustive phenomenon known as globalization. Globalization more formally refers to the economic, social, cultural, or environmental changes that tend to interconnect peoples around the world. Since the economic aspects of globalization are certainly the most pervasive of these changes, it is increasingly important to understand the implications of a global marketplace on consumers, businesses, and governments. That is where the study of international economics begins.

What Is International Economics?

International economics is a field of study that assesses the implications of international trade, international investment, and international borrowing and lending. There are two broad subfields within the discipline: international trade and international finance.

International trade is a field in economics that applies microeconomic models to help understand the international economy. Its content includes basic supply-and-demand analysis of international markets; firm and consumer behavior; perfectly competitive, oligopolistic, and monopolistic market structures; and the effects of market distortions. The typical course describes economic relationships among consumers, firms, factory owners, and the government.

The objective of an international trade course is to understand the effects of international trade on individuals and businesses and the effects of changes in trade policies and other economic conditions. The course develops arguments that support a free trade policy as well as arguments that support various types of

protectionist policies. By the end of the course, students should better understand the centuries-old controversy between free trade and protectionism.

International finance applies macroeconomic models to help understand the international economy. Its focus is on the interrelationships among aggregate economic variables such as GDP, unemployment rates, inflation rates, trade balances, exchange rates, interest rates, and so on. This field expands basic macroeconomics to include international exchanges. Its focus is on the significance of trade imbalances, the determinants of exchange rates, and the aggregate effects of government monetary and fiscal policies. The pros and cons of fixed versus floating exchange rate systems are among the important issues addressed.

This international trade textbook begins in this chapter by discussing current and past issues and controversies relating to microeconomic trends and policies. We will highlight past trends both in implementing policies that restrict trade and in forging agreements to reduce trade barriers. It is these real-world issues that make the theory of international trade worth studying.

KEY TAKEAWAYS

- International trade and investment flows have grown dramatically and consistently during the past half century.
- International trade is a field in economics that applies microeconomic models to help understand the international economy.
- International finance focuses on the interrelationships among aggregate economic variables such as GDP, unemployment, inflation, trade balances, exchange rates, and so on.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The approximate share of world exports as a percentage of world GDP in 2008.
 - b. The approximate share of world foreign direct investment as a percentage of world GDP in 1980.
 - c. The number of countries that were members of the WTO in 2009.
 - d. This branch of international economics applies microeconomic models to understand the international economy.
 - e. This branch of international economics applies macroeconomic models to understand the international economy.

1.2 Understanding Tariffs

LEARNING OBJECTIVES

1. Learn the different methods used to assess a tariff.
2. Measure, interpret, and compare average tariffs around the world.

The most common way to protect one's economy from import competition is to implement a tariff: a tax on imports. Generally speaking, a tariff is any tax or fee collected by a government. Sometimes the term "tariff" is used in a nontrade context, as in railroad tariffs. However, the term is much more commonly used to refer to a tax on imported goods.

Tariffs have been applied by countries for centuries and have been one of the most common methods used to collect revenue for governments. Largely this is because it is relatively simple to place customs officials at the border of a country and collect a fee on goods that enter. Administratively, a tariff is probably one of the easiest taxes to collect. (Of course, high tariffs may induce smuggling of goods through nontraditional entry points, but we will ignore that problem here.)

Tariffs are worth defining early in an international trade course since changes in tariffs represent the primary way in which countries either liberalize trade or protect their economies. It isn't the only way, though, since countries also implement subsidies, quotas, and other types of regulations that can affect trade flows between countries. These other methods will be defined and discussed later, but for now it suffices to understand tariffs since they still represent the basic policy affecting international trade patterns.

When people talk about trade liberalization, they generally mean reducing the tariffs on imported goods, thereby allowing the products to enter at lower cost. Since lowering the cost of trade makes it more profitable, it will make trade freer. A complete elimination of tariffs and other barriers to trade is what economists and others mean by free trade. In contrast, any increase in tariffs is referred to as protection, or protectionism. Because tariffs raise the cost of importing products from abroad but not from domestic firms, they have the effect of protecting the domestic firms that compete with imported products. These domestic firms are called import competitors.

There are two basic ways in which tariffs may be levied: specific tariffs and ad valorem tariffs. A specific tariff is levied as a fixed charge per unit of imports. For example, the U.S. government levies a \$0.51 specific tariff on every wristwatch imported into the United States. Thus, if one thousand watches are imported, the U.S. government collects \$510 in tariff revenue. In this case, \$510 is collected whether the watch is a \$40 Swatch or a \$5,000 Rolex.

An ad valorem tariff is levied as a fixed percentage of the value of the commodity imported. “Ad valorem” is Latin for “on value” or “in proportion to the value.” The United States currently levies a 2.5 percent ad valorem tariff on imported automobiles. Thus, if \$100,000 worth of automobiles are imported, the U.S. government collects \$2,500 in tariff revenue. In this case, \$2,500 is collected whether two \$50,000 BMWs or ten \$10,000 Hyundais are imported.

Occasionally, both a specific and an ad valorem tariff are levied on the same product simultaneously. This is known as a two-part tariff. For example, wristwatches imported into the United States face the \$0.51 specific tariff as well as a 6.25 percent ad valorem tariff on the case and the strap and a 5.3 percent ad valorem tariff on the battery. Perhaps this should be called a three-part tariff!

As the above examples suggest, different tariffs are generally applied to different commodities. Governments rarely apply the same tariff to all goods and services imported into the country. Several countries prove the exception, though. For example, Chile levies a 6 percent tariff on every imported good, regardless of the category. Similarly, the United Arab Emirates sets a 5 percent tariff on almost all items, while Bolivia levies tariffs either at 0 percent, 2.5 percent, 5 percent, 7.5 percent, or 10 percent. Nonetheless, simple and constant tariffs such as these are uncommon.

Thus, instead of one tariff rate, countries have a tariff schedule that specifies the tariff collected on every particular good and service. In the United States, the tariff schedule is called the Harmonized Tariff Schedule (HTS) of the United States. The commodity classifications are based on the international Harmonized Commodity Coding and Classification System (or the Harmonized System) established by the World Customs Organization.

Tariff rates for selected products in the United States in 2009 are available in [Chapter 1 "Introductory Trade Issues: History, Institutions, and Legal Framework", Section 1.8 "Appendix A: Selected U.S. Tariffs—2009".](#)

Measuring Protectionism: Average Tariff Rates around the World

One method used to measure the degree of protectionism within an economy is the average tariff rate. Since tariffs generally reduce imports of foreign products, the higher the tariff, the greater the protection afforded to the country's import-competing industries. At one time, tariffs were perhaps the most commonly applied trade policy. Many countries used tariffs as a primary source of funds for their government budgets. However, as trade liberalization advanced in the second half of the twentieth century, many other types of nontariff barriers became more prominent.

[Table 1.1 "Average Tariffs in Selected Countries \(2009\)"](#) provides a list of average tariff rates in selected countries around the world. These rates were calculated as the simple average tariff across more than five thousand product categories in each country's applied tariff schedule located on the World Trade Organization (WTO) Web site. The countries are ordered by highest to lowest per capita income.

Table 1.1 Average Tariffs in Selected Countries (2009)

Country	Average Tariff Rates (%)
United States	3.6
Canada	3.6
European Community (EC)	4.3
Japan	3.1
South Korea	11.3
Mexico	12.5
Chile	6.0 (uniform)
Argentina	11.2
Brazil	13.6
Thailand	9.1
China	9.95
Egypt	17.0
Philippines	6.3
India	15.0
Kenya	12.7

Country	Average Tariff Rates (%)
Ghana	13.1

Generally speaking, average tariff rates are less than 20 percent in most countries, although they are often quite a bit higher for agricultural commodities. In the most developed countries, average tariffs are less than 10 percent and often less than 5 percent. On average, less-developed countries maintain higher tariff barriers, but many countries that have recently joined the WTO have reduced their tariffs substantially to gain entry.

Problems Using Average Tariffs as a Measure of Protection

The first problem with using average tariffs as a measure of protection in a country is that there are several different ways to calculate an average tariff rate, and each method can give a very different impression about the level of protection.

The tariffs in [Table 1.1 "Average Tariffs in Selected Countries \(2009\)"](#) are calculated as a simple average. To calculate this rate, one simply adds up all the tariff rates and divides by the number of import categories. One problem with this method arises if a country has most of its trade in a few categories with zero tariffs but has high tariffs in many categories it would never find advantageous to import. In this case, the average tariff may overstate the degree of protection in the economy.

This problem can be avoided, to a certain extent, if one calculates the trade-weighted average tariff. This measure weighs each tariff by the share of total imports in that import category. Thus, if a country has most of its imports in a category with very low tariffs but has many import categories with high tariffs and virtually no imports, then the trade-weighted average tariff would indicate a low level of protection. The simple way to calculate a trade-weighted average tariff rate is to divide the total tariff revenue by the total value of imports. Since these data are regularly reported by many countries, this is a common way to report average tariffs. To illustrate the difference, the United States is listed in [Table 1.1 "Average Tariffs in Selected Countries \(2009\)"](#) with a simple average tariff of 3.6 percent. However, in 2008 the U.S. tariff revenue collected came to \$29.2 billion from imports of goods totaling \$2,126 billion, meaning that the U.S. trade-weighted average tariff was a mere 1.4 percent.

Nonetheless, the trade-weighted average tariff is not without flaws. For example, suppose a country has relatively little trade because it has prohibitive tariffs (i.e., tariffs set so high as to eliminate imports) in many import categories. If it has some trade in a few import categories with relatively low tariffs, then the trade-weighted

average tariff would be relatively low. After all, there would be no tariff revenue in the categories with prohibitive tariffs. In this case, a low average tariff could be reported for a highly protectionist country. Also, in this case, the simple average tariff would register as a higher average tariff and might be a better indicator of the level of protection in the economy.

Of course, the best way to overstate the degree of protection is to use the average tariff rate on *dutiable* imports. This alternative measure, which is sometimes reported, only considers categories in which a tariff is actually levied and ignores all categories in which the tariff is set to zero. Since many countries today have many categories of goods with zero tariffs applied, this measure would give a higher estimate of average tariffs than most of the other measures.

The second major problem with using average tariff rates to measure the degree of protection is that tariffs are not the only trade policy used by countries. Countries also implement quotas, import licenses, voluntary export restraints, export taxes, export subsidies, government procurement policies, domestic content rules, and much more. In addition, there are a variety of domestic regulations that, for large economies at least, can and do have an impact on trade flows. None of these regulations, restrictions, or impediments to trade, affecting both imports and exports, would be captured using any of the average tariff measures. Nevertheless, these nontariff barriers can have a much greater effect on trade flows than tariffs themselves.

KEY TAKEAWAYS

- Specific tariffs are assessed as a money charge per unit of the imported good.
- Ad valorem tariffs are assessed as a percentage of the value of the imported good.
- Average tariffs can be measured as a simple average across product categories or can be weighted by the level of imports.
- Although average tariffs are used to measure the degree of protection or openness of a country, neither measure is best because each measure has unique problems.
- In general, average tariffs are higher in developing countries and lower in developed countries.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. A type of tariff assessed as a percentage of the value of the imported good (e.g., 12 percent of the value of apples).
 - b. A type of tariff assessed as a fixed money charge per unit of imports (e.g., \$0.35 per pound of apples).
 - c. Of *increase* or *decrease*, this is how tariffs would be changed if a country is liberalizing trade.
2. Calculate the amount of tariff revenue collected if a 7 percent ad valorem tariff is assessed on ten auto imports with the autos valued at \$20,000 each.
3. Calculate the amount of tariff revenue collected if a \$500 specific tariff is assessed on ten auto imports with the autos valued at \$20,000 each.
 - a. What would the ad valorem tariff rate have to be to collect the same amount of tariff revenue?
4. Calculate the trade-weighted average tariff if a country has annual goods imports of \$157 billion and annual tariff revenue of \$13.7 billion.

1.3 Recent Trade Controversies

LEARNING OBJECTIVES

1. Identify some of the ways the world has stepped closer to free trade recently.
2. Identify some of the ways the world has stepped further from free trade recently.

In the spring of 2009, the world was in the midst of the largest economic downturn since the early 1980s. Economic production was falling and unemployment was rising. International trade had fallen substantially everywhere in the world, while investment both domestically and internationally dried up.

The source of these problems was the bursting of a real estate bubble. Bubbles are fairly common in both real estate and stock markets. A bubble describes a steady and persistent increase in prices in a market—in this case, in the real estate markets in the United States and abroad. When bubbles are developing, many market observers argue that the prices are reflective of true values despite a sharp and unexpected increase. These justifications fool many people into buying the products in the hope that the prices will continue to rise and generate a profit.

When the bubble bursts, the demand driving the price increases ceases and a large number of participants begin to sell off their product to realize their profit. When this occurs, prices quickly plummet. The dramatic drop in real estate prices in the United States in 2007 and 2008 left many financial institutions near bankruptcy. These financial market instabilities finally spilled over into the *real sector* (i.e., the sector where goods and services are produced), contributing not only to a world recession but also to a new popular attitude that capitalism and free markets may not be working very well. This attitude change may fuel the antiglobalization sentiments that were growing during the previous decade.

As the current economic crisis unfolded, there were numerous suggestions about similarities between this recession and the Great Depression in the 1930s. One big concern was that countries might revert to protectionism to try to save jobs for domestic workers. This is precisely what many countries did at the onset of the Great Depression, and it is widely believed that that reaction made the Depression worse rather than better.

Since the economic crisis began in late 2008, national leaders have regularly vowed to avoid protectionist pressures and maintain current trade liberalization commitments made under the World Trade Organization (WTO) and individual free trade agreements. However, at the same time, countries have raised barriers to trade in a variety of subtle ways. For example, the United States revoked a promise to maintain a program allowing Mexican trucks to enter the United States under the North American Free Trade Agreement (NAFTA), it included “Buy American” provisions in its economic stimulus package, it initiated a special safeguards action against Chinese tire imports, and it brought a case against China at the WTO. Although many of these actions are legal and allowable under U.S. international commitments, they are nevertheless irritating to U.S. trading partners and indicative of the rising pressure to implement policies favorable to domestic businesses and workers. Most other countries have taken similar, albeit subtle, protectionist actions as well.

Nevertheless, this rising protectionism runs counter to a second popular sentiment among people seeking to achieve greater liberalization and openness in international markets. For example, as the recession began, the United States had several free trade areas waiting to be approved by the U.S. Congress: one with South Korea, another with Colombia, and a third with Panama. In addition, the United States has participated in talks recently with many Pacific Rim countries to forge a Trans-Pacific Partnership (TPP) that could liberalize trade around the region. Simultaneously, free trade area discussions continue among many other country pairings around the world.

This current ambivalence among countries and policymakers is nothing new. Since the Great Depression, trade policymaking around the world can be seen as a tug of war between proponents and opponents of trade liberalization. Even as free trade advocates have achieved trade expansions and liberalizations, free trade opponents have often achieved market-closing policies at the same time; three steps forward toward trade liberalization are often coupled with two steps back at the same time.

To illustrate this point, we continue with a discussion of both recent initiatives for trade liberalization and some of the efforts to resist these liberalization movements. We'll also look back to see how the current policies and discussions have been shaped by events in the past century.

Doha and WTO

The Doha Round is the name of the current round of trade liberalization negotiations undertaken by WTO member countries. The objective is for all participating countries to reduce trade barriers from their present levels for trade

in goods, services, and agricultural products; to promote international investment; and to protect intellectual property rights. In addition, member countries discuss improvements in procedures that outline the rights and responsibilities of the member countries. Member countries decided that a final agreement should place special emphasis on changes targeting the needs of developing countries and the world's poor and disadvantaged. As a result, the Doha Round is sometimes called the Doha Development Agenda, or DDA.

The Doha Round was begun at the WTO ministerial meeting held in Doha, Qatar, in November 2001. It is the first round of trade liberalization talks under the auspices of the WTO, which was founded in 1994 in the final General Agreement on Tariffs and Trade (GATT) round of talks, the Uruguay Round. Because missed deadlines are commonplace in the history of GATT talks, an old joke is that GATT really means the “General Agreement to Talk and Talk.”

In anticipation, WTO members decided to place strict deadlines for different phases of the agreement. By adhering to the deadlines, countries were more assured that the talks would be completed on schedule in the summer of 2005—but the talks weren’t. So members pushed off the deadline to 2006, and then to 2007, and then to 2008, always reporting that an agreement was near. As of 2009, the Doha Round has still not been completed, testifying to the difficulty of getting 153 member countries to conceive of a trade liberalization agreement that all countries can accept mutually.

This is an important point: WTO rounds (and the GATT rounds before them) are never finalized until every member country agrees to the terms and conditions. Each country offers a set of trade-liberalizing commitments, or promises, and in return receives the trade-liberalizing commitments made by its 152 potential trading partners. This is a much stronger requirement than majority voting, wherein coalitions can force other members into undesirable outcomes. Thus one reason this round has so far failed is because some countries believe that the others are offering too little liberalization relative to the liberalization they themselves are offering.

The DDA is especially complex, not only because 153 countries must reach a consensus, but also because there are so many trade-related issues under discussion. Countries discuss not only tariff reductions on manufactured goods but also changes in agricultural support programs, regulations affecting services trade, intellectual property rights policy and enforcement, and procedures involving trade remedy laws, to name just a few. Reaching an agreement that every country is happy about across all these issues may be more than the system can handle. We’ll have to wait to see whether the Doha Round ever finishes to know if it is possible.

Even then, there is some chance an agreement that is achievable may be so watered down that it doesn't result in much trade liberalization.

The primary stumbling block in the Doha Round (and the previous Uruguay Round too) has been insufficient commitments on agricultural liberalization, especially by the developed countries. Today, agriculture remains the most heavily protected industry around the world. In addition to high tariffs at the borders, most countries offer subsidies to farmers and dairy producers, all of which affects world prices and international trade. Developing countries believe that the low world prices for farm products caused by subsidies in rich countries both prevents them from realizing their comparative advantages and stymies economic development. However, convincing developed country farmers to give up long-standing handouts from their governments has been a difficult to impossible endeavor.

To their credit, developed countries have suggested that they may be willing to accept greater reductions in agricultural subsidies if developing countries would substantially reduce their very high tariff bindings on imported goods and bind most or all of their imported products. Developing countries have argued, however, that because this is the Doha "Development" Round, they shouldn't be asked to make many changes at all to their trade policies; rather, they argue that changes should be tilted toward greater market access from developing into developed country markets.

Of course, this is not the only impasse in the discussions, as there are many other issues on the agenda. Nevertheless, agricultural liberalization will surely remain one of the major stumbling blocks to continued trade liberalization efforts. And the Doha Round is not dead yet, since continuing discussions behind the spotlight reflect at least some sentiment around the world that further trade liberalization is a worthy goal. But this is not a sentiment shared by all, and indeed opponents almost prevented this WTO round from beginning in the first place. To understand why, we need to go back two years to the Doha Round commencement in Seattle, Washington, in December 1999.

The WTO Seattle Ministerial—1999

Every two years, the WTO members agreed to hold a ministerial meeting bringing together, at minimum, the trade ministers of the member countries to discuss WTO issues. In 1999, the ministerial was held in Seattle, Washington, in the United States, and because it was over five years since the last round of trade discussions had finished, many members thought it was time to begin a new round of trade talks. There is a well-known "bicycle theory" about international trade talks that says

that forward momentum must be maintained or else, like a bicycle, liberalization efforts will stall.

And so the WTO countries decided by 1999 to begin a new “Millennial Round” of trade liberalization talks and to kick off the discussions in Seattle in December 1999. However, two things happened, the first attesting to the difficulty of getting agreement among so many countries and the second attesting to the growing opposition to the principles of free trade itself.

Shortly before the ministers met, they realized that there was not even sufficient agreement among governments about what the countries should discuss in the new round. For example, the United States was opposed to any discussion about trade remedy laws, whereas many developing countries were eager to discuss revisions. Consequently, because no agreement—even about what to talk about—could be reached, the start of the round was postponed.

The second result of the meeting was a cacophony of complaints that rose up from the thousands of protesters who gathered outside the meetings. This result was more profound if only because the resulting disturbances, including property damage and numerous arrests, brought the issues of trade and the WTO to the international stage. Suddenly, the world saw that there was substantial opposition to the principles of the WTO in promoting trade and expanded globalization.

These protests at the Seattle Ministerial were perhaps directed not solely at the WTO itself but instead at a variety of issues brought to the forefront by globalization. Some protesters were there to protest environmental degradation and were worried that current development was unsustainable, others were protesting child labor and unsafe working conditions in developing countries, and still others were concerned about the loss of domestic jobs due to international competition. In many ways, the protesters were an eclectic group consisting of students, labor union members, environmentalists, and even some anarchists.

After Seattle, groups sometimes labeled “antiglobalization groups” began organizing protests at other prominent international governmental meetings, including the biannual World Bank and International Monetary Fund (IMF) meetings, the meeting of the G8 countries, and the World Economic Forum at Davos, Switzerland. The opposition to freer trade, and globalization more generally, was on the rise. At the same time, though, national governments continued to press for more international trade and investment through other means.

Ambivalence about Globalization since the Uruguay Round

Objectively speaking, ambivalence about trade and globalization seems to best characterize the decades of the 1990s and 2000s. Although this was a time of rising protests and opposition to globalization, it was also a time in which substantial movements to freer trade occurred. What follows are some events of the last few decades highlighting this ambivalence.

First off, trade liberalization became all the rage around the world by the late 1980s. The remarkable success of outward-oriented economies such as South Korea, Taiwan, Hong Kong, and Singapore—known collectively as the East Asian Tigers—combined with the relatively poor performance of inward-oriented economies in Latin America, Africa, India, and elsewhere led to a resurgence of support for trade.

Because the Uruguay Round of the GATT was on its way to creating the WTO, many countries decided to jump on the liberalizing bandwagon by joining the negotiations to become founding members of the WTO. One hundred twenty-three countries were members of the WTO upon its inception in 1995, only to grow to 153 members by 2009.

Perhaps the most important new entrant into the WTO was China in 2001. China had wanted to be a founding member of the WTO in 1995 but was unable to overcome the accession hurdle. You see, any country that is already a WTO member has the right to demand trade liberalization concessions from newly acceding members. Since producers around the world were fearful of competition from China, most countries demanded more stringent liberalization commitments than were usually expected from other acceding countries at a similar level of economic development. As a result, it took longer for China to gain entry than for most other countries.

But at the same time that many developing countries were eager to join the WTO, beliefs in freer trade and the WTO were reversing in the United States. Perhaps the best example was the struggle for the U.S. president to secure trade-negotiating authority. First, a little history.

Article 1, section 8 of the U.S. Constitution states, “The Congress shall have the power...to regulate commerce with foreign nations.” This means that decisions about trade policies must be made by the U.S. Senate and House of Representatives, and *not* by the U.S. president. Despite this, the central agency in trade negotiations today is the United States Trade Representative (USTR), an executive branch (or presidential) agency. The reason for this arrangement is that the U.S. Congress has

ceded authority for these activities to the USTR. One such piece of enabling legislation is known as trade promotion authority (TPA).

TPA enables the U.S. president, or more specifically the USTR, to negotiate trade liberalization agreements with other countries. The legislation is known as *fast-track authority* because it provides for expedited procedures in the approval process by the U.S. Congress. More specifically, for any trade agreement the president presents to the Congress, Congress will vote the agreement, in its entirety, up or down in a yea or nay vote. Congress agrees not to amend or change in any way the contents of the negotiated agreement. The fast-track procedure provides added credibility to U.S. negotiators since trade agreement partners will know the U.S. Congress cannot change the details upon review.

TPA has been given to the U.S. president in various guises since the 1930s. In the post-World War II era, authority was granted to the president to negotiate successive GATT rounds. A more recent incarnation was granted to the president in the Trade Act of 1974. TPA enabled negotiations for the U.S.-Israel free trade area (FTA) in 1985 and NAFTA in 1993. However, this authority expired in 1994 under President Clinton and was never reinstated during the remainder of his presidency. The failure to extend TPA signified the growing discontent, especially in the U.S. House of Representatives, with trade liberalization.

When George W. Bush became president, he wanted to push for more trade liberalization through the expansion of FTAs with regional and strategic trade partners. He managed to gain a renewal of TPA in 2001 (with passage in the House by just one vote, 216 to 215). This enabled President Bush to negotiate and implement a series of FTAs with Chile, Singapore, Australia, Morocco, Jordan, Bahrain, Oman, Central America and the Dominican Republic, and Peru. Awaiting congressional approval (as of December 2009) are FTAs with South Korea, Colombia, and Panama.

Despite these advances toward trade liberalization, TPA expired in 2007 and has not yet been renewed by the U.S. Congress, again representing the ambivalence of U.S. policymakers to embrace freer trade. Another indication is the fact that the FTAs with South Korea, Colombia, and Panama were submitted for approval to Congress before the deadline for TPA expired in 2007 and these agreements still have not been brought forward for a vote by the U.S. Congress.

While the United States slows its advance toward freer trade, other countries around the world continue to push forward. There are new FTAs between China and the Association of Southeast Asian Nations (ASEAN) countries, Japan and the

Philippines, Thailand and Chile, Pakistan and China, and Malaysia and Sri Lanka, along with several other new pairings.

Future prospects for trade liberalization versus trade protections are quite likely to depend on the length and severity of the present economic crisis. If the crisis abates soon, trade liberalization may return to its past prominence. However, if the crisis continues for several more years and if unemployment rates remain much higher than usual for an extended time, then demands for more trade protection may increase significantly. Economic crises have proved in the past to be a major contributor to high levels of protection. Indeed, as was mentioned previously, there is keen awareness today that the world may stumble into the trade policy mistakes of the Great Depression. Much of the trade liberalization that has occurred since then can be traced to the desire to reverse the effects of the Smoot-Hawley Tariff Act of 1930. Thus to better understand the current references to our past history, the story of the Great Depression is told next.

KEY TAKEAWAYS

- Recent support for trade liberalization is seen in the establishment of numerous free trade areas and the participation of many countries in the Doha Round of trade talks.
- Recent opposition to trade liberalization is seen in national responses to the financial crisis, the protest movement at the Seattle Ministerial and other venues, and the failure in the United States to grant trade promotion authority to the president.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This branch of the U.S. government is given the authority to make trade policy.
 - b. This theory suggests why continual negotiations are needed to assure long-term progress toward trade liberalization.
 - c. This WTO ministerial meeting in 1999 began a wave of protests around the world against globalization initiatives.
 - d. The term used to describe the U.S. presidential authority that includes expedited approval procedures in the U.S. Congress.
 - e. The names of three countries with which the United States has implemented free trade areas.
 - f. The name of the WTO round of trade liberalization talks begun in 2001.
 - g. The term used to describe the economic sector in which goods and services are produced and traded, in contrast to the monetary sector.

1.4 The Great Depression, Smoot-Hawley, and the Reciprocal Trade Agreements Act (RTAA)

LEARNING OBJECTIVE

1. Understand the trade policy effects of the Great Depression.

Perhaps the greatest historical motivator for trade liberalization since World War II was the experience of the Great Depression. The Depression ostensibly began with the crash of the U.S. stock market in late 1929. Quite rapidly thereafter, the world economy began to shrink at an alarming pace. In 1930, the U.S. economy shrank by 8.6 percent and the unemployment rate rose to 8.9 percent. With the contraction came a chorus of calls for protection of domestic industries facing competition from imported products.

For U.S. workers, a tariff bill to substantially raise protection was already working its way through the legislature when the economic crisis hit. The objective of higher tariffs was to increase the cost of imported goods so that U.S. consumers would spend their money on U.S. products instead. By doing so, U.S. jobs could be saved in the import-competing industries. Many economists at the time disagreed with this analysis and thought the high tariffs would make things worse. In May 1930, 1,028 economists signed a petition protesting the tariff act and beseeched President Hoover to veto the bill. Despite these objections, in June of 1930 the Smoot-Hawley Tariff Act (aka the Tariff Act of 1930), which raised average tariffs to as much as 60 percent, was passed into law.

However, because higher U.S. tariffs also injured the foreign companies that were exporting into the U.S. market and because the foreign economies were also stagnating and suffering from rising unemployment, they responded to the Smoot-Hawley tariffs with higher tariffs of their own in retaliation. Within several months, numerous U.S. trade partners responded by protecting their own domestic industries with higher trade barriers. The effect was a dramatic drop in international trade flows throughout the world and quite possibly a deepening of the economic crisis.

In subsequent years, the Depression did get much worse. The U.S. economy continued to contract at double-digit rates for several more years, and the unemployment rate peaked in 1933 at 24.9 percent. When Franklin Roosevelt ran

for president in 1932, he spoke against the high tariffs. By 1934, a new attitude accepting the advantages of more liberal trade took hold in the U.S. Congress, which passed the Reciprocal Trade Agreements Act (RTAA). The RTAA authorized the U.S. president to negotiate bilateral tariff reduction agreements with other countries.

In practice, the president could send his agents to another country, say Mexico, to offer tariff reductions on a collection of imported items in return for tariff reductions by Mexico on another set of items imported from the United States. Once both sides agreed to the quid pro quo, the agreements would be brought back to the United States and the Mexican governments for approval and passage into law. Over sixty bilateral deals were negotiated under the RTAA, and it set in motion a process of trade liberalization that would continue for decades to come.

The RTAA is significant for two reasons. First, it was one of the earliest times when the U.S. Congress granted trade policymaking authority directly to the president. In later years, this practice continued with congressional approval for presidential trade promotion authority (TPA; aka fast-track authority) that was used to negotiate other trade liberalization agreements. Second, the RTAA served as a model for the negotiating framework of the General Agreement on Tariffs and Trade (GATT). Under the GATT, countries would also offer “concessions,” meaning tariff reductions on imports, in return for comparable concessions from the other GATT members. The main difference is that the RTAA involved bilateral concessions, whereas the GATT was negotiated in a multilateral environment. More on the GATT next.

KEY TAKEAWAYS

- The Great Depression inspired a great wave of protectionism around the world beginning with the Smoot-Hawley Tariff Act in the United States in 1930.
- The Reciprocal Trade Agreements Act (RTAA) was the start of a wave of trade liberalization.
- The RTAA was important because it gave trade policymaking authority to the U.S. president and because it served as a model for the GATT.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The *common name* given to the U.S. Tariff Act of 1930.
 - b. The term used to describe the U.S. presidential authority to negotiate free trade areas.
 - c. The name of the 1934 U.S. legislative act that authorized the U.S. president to negotiate bilateral tariff reduction agreements.
 - d. The highest U.S. unemployment rate during the Great Depression.
 - e. The name of the U.S. president who signed the Tariff Act of 1930.
 - f. The number of economists who signed a petition protesting the Smoot-Hawley Tariff Act.

1.5 The General Agreement on Tariffs and Trade (GATT)

LEARNING OBJECTIVES

1. Learn the basic principles underpinning the GATT.
2. Identify the special provisions and allowable exceptions to the basic principles of the GATT.

The General Agreement on Tariffs and Trade (GATT) was never designed to be a stand-alone agreement. Instead, it was meant to be just one part of a much broader agreement to establish an International Trade Organization (ITO). The ITO was intended to promote trade liberalization by establishing guidelines or rules that member countries would agree to adopt. The ITO was conceived during the Bretton Woods conference attended by the main allied countries in New Hampshire in 1944 and was seen as complementary to two other organizations also conceived there: the International Monetary Fund (IMF) and the World Bank. The IMF would monitor and regulate the international fixed exchange rate system, the World Bank would assist with loans for reconstruction and development, and the ITO would regulate international trade.

The ITO never came into existence, however. Although a charter was drawn, the U.S. Congress never approved it. The main concern was that the agreement would force unwelcome domestic policy changes, especially with respect to wage and employment policies. Because the United States would not participate, other countries had little incentive to participate. Nonetheless, the United States, Britain, and other allied countries maintained a strong commitment to the reduction of tariffs on manufactured goods. Tariffs still remained high in the aftermath of the Depression-era increases. Thus, as discussions over the ITO charter proceeded, the GATT component was finalized early and signed by twenty-three countries in 1948 as a way of jump-starting the trade liberalization process.

The GATT consists of a set of promises, or commitments, that countries make to each other regarding their own trade policies. The goal of the GATT is to make trade freer (i.e., to promote trade liberalization), and thus the promises countries make must involve reductions in trade barriers. Countries that make these commitments and sign on to the agreement are called signatory countries. The discussions held before the commitments are decided are called negotiating rounds. Each round is generally given a name tied either to the location of the meetings or to a prominent figure. There were eight rounds of negotiation under the GATT: the Geneva Round

(1948), the Annecy Round (1950), the Torquay Round (1951), the Geneva II Round (1956), the Dillon Round (1962), the Kennedy Round (1967), the Tokyo Round (1979), and the Uruguay Round (1994). Most importantly, the agreements are reached by consensus. A round finishes only when every negotiating country is satisfied with the promises it and all of its negotiating partners are making. The slogan sometimes used is “Nothing Is Agreed Until Everything Is Agreed.”

The promises, or commitments, countries make under the GATT take two forms. First, there are country-specific and product-specific promises. For example, a country (say, the United States) may agree to reduce the maximum tariff charged on a particular item (say, refrigerator imports) to a particular percentage (say, 10 percent). This maximum rate is called a tariff binding, or a bound tariff rate.

In each round, every participating country offers concessions, which involve a list of new tariff bindings—one for every imported product. To achieve trade liberalization, the tariff bindings must be lower than they were previously. However, it is important to note that there is no harmonization of tariff bindings. At the end of a round, signatory countries do not end up with the same tariff rates.

Instead, each country enters a round with a unique tariff set on every item. The expectation in the negotiating round is that each country will ratchet its tariffs downward, on average, from its initial levels. Thus, if Country A enters the discussions with a 10 percent tariff on refrigerator imports, while Country B has a 50 percent tariff, then a typical outcome to the round may have A lowering its tariff binding to 7 percent, while B lowers its to 35 percent—both 30 percent reductions in the tariff binding. Both countries have liberalized trade, but the GATT has not required them to adhere to the same trade policies.

Some countries, especially developing countries, maintain fairly high bound tariffs but have decided to reduce the actual tariff to a level below the bound rate. This tariff is called the applied tariff. Lowering tariffs unilaterally is allowable under the GATT, as is raising the applied rate up to the bound rate. Further discussion of this issue can be found in [Chapter 1 "Introductory Trade Issues: History, Institutions, and Legal Framework"](#), [Section 1.9 "Appendix B: Bound versus Applied Tariffs"](#).

There is a second form of promise that GATT countries make that is harmonized. These promises involve acceptance of certain principles of behavior with respect to international trade policies. Here, too, there are two types of promises: the first involves core principles regarding nondiscrimination and the second involves allowable exceptions to these principles.

Nondiscrimination

One of the key principles of the GATT, one that signatory countries agree to adhere to, is the nondiscriminatory treatment of traded goods. This means countries assure that their own domestic regulations will not affect one country's goods more or less favorably than another country's and will not treat their own goods more favorably than imported goods. There are two applications of nondiscrimination: most-favored nation and national treatment.

Most-Favored Nation

Most-favored nation (MFN)⁴ refers to the nondiscriminatory treatment toward identical or highly substitutable goods coming from two different countries. For example, if the United States applies a tariff of 2.6 percent on printing press imports from the European Union (EU, one World Trade Organization [WTO] country), then it must apply a 2.6 percent tariff on printing press imports from every other WTO member country. Since all the countries must be treated *identically*, MFN is a bit of a misnomer since it seems to suggest that one country is *most* favored, whereas in actuality, it means that countries are *equally* favored.

The confusion the term generates led the United States in the 1990s to adopt an alternative phrase, *normal trade relations* (NTR), for use in domestic legislation. This term is a better description of what the country is offering when a new country enters the WTO or when a non-WTO country is offered the same tariff rates as its WTO partner countries. As such, these are two ways to describe the same thing: that is, MFN ≡ NTR.

National Treatment

National treatment⁵ refers to the nondiscriminatory treatment of identical or highly substitutable domestically produced goods with foreign goods once the foreign products have cleared customs. Thus it is allowable to discriminate by applying a tariff on imported goods that would not be applied to domestic goods, but once the product has passed through customs it must be treated identically. This norm applies then to both state and local taxes, as well as regulations such as those involving health and safety standards. For example, if a state or provincial government applies a tax on cigarettes, then national treatment requires that the same tax rate be applied equally on domestic and foreign cigarettes. Similarly, national treatment would prevent a government from regulating lead-painted imported toys to be sold but not lead-painted domestic toys; if lead is to be regulated, then all toys must be treated the same.

4. The nondiscriminatory treatment toward identical or highly substitutable goods coming from two different countries.
5. The nondiscriminatory treatment of identical or highly substitutable domestically produced goods with foreign goods once the foreign products have cleared customs.

GATT Exceptions

There are several situations in which countries are allowed to violate GATT nondiscrimination principles and previous commitments such as tariff bindings. These represent allowable exceptions that, when implemented according to the guidelines, are GATT sanctioned or GATT legal. The most important exceptions are trade remedies and free trade area allowances.

Trade Remedies

An important class of exceptions is known as trade remedies. These are laws that enable domestic industries to request increases in import tariffs that are above the bound rates and are applied in a discriminatory fashion. They are called remedies because they are intended to correct for unfair trade practices and unexpected changes in trade patterns that are damaging to those industries that compete with imports.

These remedies are in the GATT largely because these procedures were already a part of the laws of the United States and other allied countries when the GATT was first conceived. Since application of these laws would clearly violate the basic GATT principles of nondiscrimination, exceptions were written into the original agreement, and these remain today. As other countries have joined the GATT/WTO over the years, these countries have also adopted these same laws, since the agreement allows for them. As a result, this legal framework, established in the United States and other developed countries almost a century ago, has been exported to most other countries around the world and has become the basic method of altering trade policies from the commitments made in previous GATT rounds.

Today, the trade remedy laws represent the primary legal method WTO countries can use to raise their levels of protection for domestic industries. By binding countries to maximum levels of protection, the GATT and WTO agreements eliminate their national sovereignty with respect to higher trade barriers. Note that countries are always free to lower trade barriers unilaterally if they wish without violating the agreements. The trade remedy laws offer a kind of safety valve, because in certain prescribed circumstances, countries can essentially renege on their promises.

6. Laws that provide protection to domestic import-competing firms that can show that foreign imported products are being “dumped” in the domestic market.

Antidumping

Antidumping laws⁶ provide protection to domestic import-competing firms that can show that foreign imported products are being “dumped” in the domestic

market. Since dumping is often considered an unfair trade practice, antidumping is known as an unfair trade law. Dumping is defined in several different ways. In general, dumping means selling a product at an unfair, or less than reasonable, price. More specifically, dumping is defined as (1) sales in a foreign market at a price less than in the home market, (2) sales in a foreign market at a price that is less than average production costs, or (3) if sales in the home market do not exist, sales in one foreign market at a price that is less than the price charged in another foreign market. The percentage by which the actual price must be raised to reach the fair or reasonable price is called the dumping margin. For example, if a firm sells its product in its home market for \$12 but sells it in a foreign market for \$10, then the dumping margin is 20 percent since a 20 percent increase in the \$10 price will raise it to \$12.

Any import-competing industry is allowed to petition its own government for protection under its antidumping law. Protection in the form of an antidumping (AD) duty (i.e., a tariff on imports) can be provided if two conditions are satisfied. First, the government must show that dumping, as defined above, is actually occurring. Second, the government must show that the import-competing firms are suffering from, or are threatened with, material injury as a result of the dumped imports. Injury might involve a reduction in revenues, a loss of profit, declining employment, or other indicators of diminished well-being. If both conditions are satisfied, then an AD duty set equal to the dumping margin can be implemented. After the Uruguay Round, countries agreed that AD duties should remain in place for no more than five years before a review (called a *sunset review*) must be conducted to determine if the dumping is likely to recur. If a recurrence of dumping is likely, the AD duties may be extended.

Normally, AD investigations determine different dumping margins, even for different firms from the same country. When AD duties are applied, these different firms will have separate tariffs applied to their products. Thus the action is highly discriminatory and would normally violate MFN treatment. The increase in the tariff would also raise it above the bound tariff rate the country reached in the latest negotiating round. However, Article 6 of the original GATT allows this exception.

Antisubsidy

7. Laws that provide protection to domestic import-competing firms that can show that foreign imported products are being directly subsidized by the foreign government.

Antisubsidy laws⁷ provide protection to domestic import-competing firms that can show that foreign imported products are being directly subsidized by the foreign government. Since foreign subsidies are considered an unfair trade practice, antisubsidy is considered an unfair trade law. The subsidies must be ones that are targeted at the export of a particular product. These are known as *specific subsidies*. In contrast, *generally available subsidies*, those that apply to both export firms and

domestic firms equally, are not actionable under this provision. The percentage of the subsidy provided by the government is known as the subsidy margin.

Import-competing firms have two recourses in the face of a foreign government subsidy. First, they can appeal directly to the WTO using the dispute settlement procedure (described in [Chapter 1 "Introductory Trade Issues: History, Institutions, and Legal Framework"](#), [Section 1.7 "The World Trade Organization"](#)). Second, they can petition their own government under their domestic antisubsidy laws. In either case, they must demonstrate two things: (1) that a subsidy is being provided by the foreign government and (2) that the resulting imports have caused injury to the import-competing firms. If both conditions are satisfied, then a country may implement a *countervailing duty* (CVD)—that is, a tariff on imports set equal to the subsidy margin. As with AD duties, CVDs should remain in place for no more than five years before a sunset review must be conducted to determine if the subsidies continue. If they are still in place, the CVD may be extended.

Since CVDs are generally applied against one country's firms but not another's, the action is discriminatory and would normally violate MFN treatment. The higher tariff would also raise it above the bound tariff rate the country reached in the latest negotiating round. Nonetheless, Article 6 of the original GATT allows this exception.

Safeguards

Safeguard laws (aka escape clauses)⁸ provide protection to domestic import-competing firms that can demonstrate two things: (1) that a surge of imported products has caused disruption in the market for a particular product and (2) that the surge has substantially caused, or threatens to cause, serious injury to the domestic import-competing firms. The use of the term *serious injury* means that the injury must be more severe than the injury cause in AD and antisubsidy cases. Since import surges are not generally considered to be under the control of the exporting firms or government, safeguard laws are not considered unfair trade laws.

In the event both conditions are satisfied, a country may respond by implementing either tariffs or quotas to protect its domestic industry. If tariffs are used, they are to be implemented in a nondiscriminatory fashion, meaning they are executed equally against all countries. However, if quotas are used, they may be allocated in a way that favors some trading partners more than others. Safeguard actions are also intended to be temporary, lasting no more than four years.

8. Laws that provide protection to domestic import-competing firms that suffer a surge of imports.

As with antidumping and antisubsidy cases, because a safeguard response involves higher levels of protection, it will likely conflict with the previously agreed bound

tariff rates and thus violate the GATT principles. However, Article 19 of the GATT, the so-called *escape clause*, provides for an exception to the general rules in this case.

Because safeguard actions in effect take away some of the concessions a country has made to others, countries are supposed to give something back in return. An example of acceptable compensation would be the reduction of tariffs on some other items. This extra requirement, together with the need to establish serious rather than material injury, have contributed to making the use of safeguard actions less common relative to antidumping and antisubsidy actions.

China's Special Safeguards. When China was accepted as a WTO member country in 2001, it agreed to many demands made by other WTO members. One such provision requested by the United States was allowance for a “special safeguard provision.” The agreement reached allowed the United States and all other WTO countries to implement additional safeguard provisions on specific products from China that might suddenly flood their markets.

One important concern at the time was the surge of textile and apparel products that might come after the expiration of the quota system in 2005 under the Uruguay Round’s Agreement on Textiles and Clothing. As a stopgap, countries were allowed to reintroduce quotas or other barriers in the event that imports from China surged in once the official quotas were gone. Both the United States and the EU implemented increased protections in 2005, and China did not enjoy the full benefit of the quota elimination until this safeguard provision expired in 2008.

Additional special safeguards are in place to protect against import surges of other products from China, and these do not expire until 2014. (In the United States, these are called section 421 cases.) Although these provisions are similar to the standard safeguards, they are more lenient in defining an actionable event.

Free Trade Areas

One other common situation requires an exception to the rules of the GATT/WTO. Many countries have decided to take multiple paths toward trade liberalization. The multilateral approach describes the process of the GATT, whereby many countries simultaneously reduce their trade barriers, but not to zero. The alternative approach is referred to as regionalism, whereby two to several countries agree to reduce their tariffs and other barriers to zero—but only among themselves.

This is called a regional approach since most times the free trade partners are nearby, or at the very least are significant trading partners (though this isn't always the case).

In principle, a free trade agreement means free trade will be implemented on all products traded between the countries. In practice, free trade areas often fall short. First, they are rarely implemented immediately; instead, they are put into place over a time horizon of ten, fifteen, or even twenty or more years. Thus many free trade areas (FTAs) today are really in transition to freer trade. Second, FTAs sometimes exempt some products from liberalization. This occurs because of strong political pressure by some domestic industries. If a substantial number of products are exempted, the area is known as a preferential trade arrangement, or a PTA.

Perhaps the most important free trade area implemented in the past fifty years was the European Economic Community formed by the major countries in Western Europe in 1960 that ultimately led to the formation of the European Union in 1993. The term "union" refers to the fact that the area is now a customs union that not only includes free trade in goods and services but also allows for the mobility of workers and other factors of production. In addition, some of the core European countries have taken it one step further by creating and using the euro as a common currency, thus establishing a monetary union in addition to the customs union.

In the United States, an FTA was first implemented with Israel in 1986. An FTA with Canada in 1988 and the inclusion of Mexico with Canada to form the North American Free Trade Agreement (NAFTA) followed. Since the turn of the millennium, the United States has implemented FTAs with Jordan, Bahrain, Morocco, Singapore, Chile, Australia, the Central American Free Trade Agreement—Dominican Republic (CAFTA-DR), and Peru.

An FTA violates the GATT/WTO principle of most-favored nation because MFN requires countries to offer their most liberal trade policy to all GATT/WTO members. When an FTA is formed, the most liberal policy will become a zero tariff, or free trade. However, the original GATT carved out an exception to this rule by including Article 24. Article 24 allows countries to pair up and form free trade areas as long as the FTA moves countries significantly close to free trade and as long as countries notify the GATT/WTO of each new agreement. The simple logic is that an FTA is in the spirit of the GATT since it does involve trade liberalization.

As of 2009, over two hundred FTAs have been notified either to the GATT or the WTO. Many of these have been started in the past fifteen to twenty years, suggesting that regional approaches to trade liberalization have become more

popular, especially as progress in the multilateral forum has slowed. This trend has also fueled debate about the most effective way to achieve trade liberalization. For example, is the regional approach a substitute or complement to the multilateral approach?

KEY TAKEAWAYS

- The most-favored nation (MFN) principle of the GATT requires countries to provide nondiscriminatory treatment between identical or highly substitutable goods coming from two different countries.
- The national treatment principle of the GATT requires countries to provide nondiscriminatory treatment between identical or highly substitutable goods produced domestically and those imported from another country.
- Trade remedy laws such as antidumping, antisubsidy, and safeguards provide GATT-allowable exceptions to previous commitments and the fundamental principles.
- Although bilateral or regional free trade areas violate MFN, they are allowed by GATT because they are consistent with the goal of trade liberalization.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The name for a tariff used to offset the effects of a foreign government export subsidy in an antisubsidy action.
 - b. The international agreement established in 1948 designed to foster trade liberalization.
 - c. The term used to describe sales made by a foreign firm at a price determined to be less than reasonable value.
 - d. The WTO principle to provide the same treatment to imports from two separate WTO countries.
 - e. The WTO principle to treat an imported product in the same way as a domestically produced product.
 - f. The U.S. term used as a synonym for *most favored nation*.
 - g. The term used to describe laws that enable domestic industries to request increases in import tariffs that would otherwise violate WTO commitments.
 - h. The term used to describe a five-year review of a previous antidumping action.
 - i. The name for a WTO-sanctioned trade law that protects an industry from a surge of imports.
 - j. GATT Article 24 provides an exception for free trade areas because they violate this GATT principle.
2. What is an antidumping duty? How is its size determined?
 - a. What must U.S. government agencies determine before applying antidumping duties against foreign firms?
 - b. How does U.S. trade law define dumping?
3. What is a countervailing duty? How is its size determined?
 - a. What must U.S. government agencies determine before applying a countervailing duty against foreign firms?

1.6 The Uruguay Round

LEARNING OBJECTIVE

1. Learn how the Uruguay Round of the General Agreement on Tariffs and Trade (GATT) greatly expanded the coverage of trade liberalization efforts to previously uncovered sectors.

The Uruguay Round was the last of eight completed rounds of the GATT. Discussion for the round began in Montevideo, Uruguay, in 1986, and it was hoped that the round would be completed by 1990. However, impasses were frequent, and the round was not finalized until 1994. One reason for the delay is that this round incorporated many new issues in the negotiations.

In earlier rounds, the primary focus was always a continuing reduction in the bound tariff rates charged on imported manufactured goods. As a result of seven completed GATT rounds, by the mid-1980s tariffs in the main developed countries were as low as 5 percent to 10 percent and there was less and less room for further liberalization. At the same time, there were a series of trade issues that sidestepped the GATT trade liberalization efforts over the years. In those areas—like agriculture, textiles and apparel, services, and intellectual property—trade barriers of one sort or another persisted. Thus the ambitious objective of the Uruguay Round was to bring those issues to the table and try to forge a more comprehensive trade liberalization agreement. The goals were reached by establishing a series of supplementary agreements on top of the traditional tariff reduction commitments of the GATT. A few of these agreements are highlighted next.

The Agreement on Agriculture (AoA)

Protections and support for agricultural industries began wholeheartedly during the Great Depression in the 1930s. Not only were tariffs raised along with most other import products, but a series of price and income support programs were implemented in many countries. When the first GATT agreement was negotiated, special exceptions for agriculture were included, including an allowance to use export subsidies. Recall that export subsidies are subject to retaliation under the antisubsidy code but that requirement was negated for agricultural products. This enabled countries to keep prices for farm products high in the domestic market and, when those prices generated a surplus of food, to dump that surplus on international markets by using export subsidies.

The result of this set of rules implemented worldwide was a severe distortion in agricultural markets and numerous problems, especially for developing countries, whose producers would regularly be forced to compete with low-priced subsidized food for the developed world.

The intention at the start of the Uruguay Round was a major reduction in tariffs and quotas and also in domestic support programs. Indeed, in the United States, the Reagan administration initially proposed a complete elimination of all trade-distorting subsidies to be phased in over a ten-year period. What ultimately was achieved was much more modest. The Uruguay Round agreement missed its deadlines several times because of the reluctance of some countries, especially the European Community (EC), to make many concessions to reduce agricultural subsidies.

Countries did agree to one thing: to make a transition away from quota restrictions on agricultural commodity imports toward tariffs instead—a process called **tarification**⁹. The logic is that tariffs are more transparent and would be easier to negotiate downward in future World Trade Organization (WTO) rounds. A second concession countries made was to accept at least low levels of market access for important commodities. For many countries, important food products had prohibitive quotas in place. A prime example was the complete restriction on rice imports to Japan. The mechanism used to guarantee these minimum levels was to implement tariff-rate quotas. A **tariff-rate quota**¹⁰ sets a low tariff on a fixed quantity of imports and a high tariff on any imports over that quota. By setting the quota appropriately and setting a relatively low tariff on that amount, a country can easily meet its target minimum import levels.

The General Agreement on Trade in Services (GATS)

Trade in services has become an increasingly important share of international trade. Trade in transportation, insurance, banking, health, and other services now accounts for over 20 percent of world trade. However, trade in services is not restricted by tariffs, largely because services are not shipped in a container on a ship, truck, or train. Instead, they are transmitted in four distinct ways. First, they are transmitted by mail, phone, fax, or the Internet; this is called *cross-border supply* of services, or Mode 1. Second, services are delivered when foreign residents travel to a host country; this is called *consumption abroad*, or Mode 2. Third, services trade occurs when a foreign company establishes a subsidiary abroad; this is called *commercial presence*, or Mode 3. Finally, services are delivered when foreign residents travel abroad to supply them; this is called *presence of natural persons*, or Mode 4. Because of the transparent nature of services, economists often refer to services as “invisibles trade.”

9. A process of converting import quotas to import tariffs. WTO countries agreed to tarification for all commodities in the Uruguay Round Agreement.

10. a low tariff set on a fixed quota of imports and a high tariff set on any imports that occur over that quota.

Because services are delivered invisibly, services trade is affected not by tariffs but rather by domestic regulations. For example, the United States has a law in place called the Jones Act, which prohibits products being transported between two U.S. ports on a foreign ship. Consider this circumstance: a foreign ship arrives at one U.S. port and unloads half its cargo. It then proceeds to a second U.S. port where it unloads the remainder. During the trip between ports 1 and 2, the ship is half empty and the shipping company may be quite eager to sell cargo transport services to U.S. firms. After all, since the ship is going to port 2 anyway, the marginal cost of additional cargo is almost zero. This would be an example of Mode 1 services trade, except for the fact that the Jones Act prohibits this activity even though these services could be beneficial to both U.S. firms and to the foreign shipping company.

The Jones Act is only one of innumerable domestic regulations in the United States that restrict foreign supply of services. Other countries maintain numerous regulations of their own, restricting access to U.S. and other service suppliers as well. When the original GATT was negotiated in the 1940s, services trade was relatively unimportant, and thus at the time there was no discussion of services regulations affecting trade. By the time of the Uruguay Round, however, services trade was increasingly important, and yet there were no provisions to discuss regulatory changes that could liberalize services trade. The Uruguay Round changed that.

As a result of Uruguay Round negotiations, GATT member countries introduced the General Agreement on Trade in Services, or GATS. The GATS includes a set of specific commitments countries have made to each other with respect to market access, market access limitations, and exceptions to national treatment in specified services. For example, a country may commit to allowing foreign insurance companies to operate without restrictions. Alternatively, a country may specify limitations perhaps restricting foreign insurance company licenses to a fixed number. A country can also specify a national treatment exception if, say, domestic banks are to be granted certain privileges that foreign banks are not allowed.

Most importantly, if exceptions have not been specified, countries have agreed to maintain most-favored nation (MFN) and national treatment with respect to services provision. This is an important step in the direction of trade liberalization largely because a previously uncovered area of trade that is rapidly growing is now a part of the trade liberalization effort.

The Agreement on Textiles and Clothing (ATC)

During the 1950s, 1960s, and 1970s, as tariffs were being negotiated downward, another type of trade restriction was being used in the textile and apparel industry: voluntary export restraints. A voluntary export restraint (VER) is a restriction set by a government on the quantity of goods that can be exported out of a country during a specified period of time. Often the word “voluntary” is placed in quotes because these restraints were often implemented upon the insistence of the importing nations.

For example, in the mid 1950s, U.S. cotton textile producers faced increases in Japanese exports of cotton textiles that negatively affected their profitability. The U.S. government subsequently negotiated a VER on cotton textiles with Japan. Afterward, textiles began to flood the U.S. market from other sources like Taiwan and South Korea. A similar wave of imports affected the nations in Europe.

The United States and Europe responded by negotiating VERs on cotton textiles with those countries. By the early 1960s, other textile producers, who were producing clothing using the new synthetic fibers like polyester, began to experience the same problem with Japanese exports that cotton producers faced a few years earlier. So VERs were negotiated on exports of synthetic fibers, first from Japan and eventually from many other Southeast Asian nations. These bilateral VERs continued until eventually exporters and importers of textile products around the world held a multilateral negotiation resulting in the Multi-Fiber Agreement (MFA) in 1974. The MFA specified quotas on exports from all major exporting countries to all major importing countries. Essentially, it represented a complex arrangement of multilateral VERs.

The MFA was renewed periodically throughout the 1970s, 1980s, and 1990s, and it represented a significant setback in the pursuit of trade liberalization. Thus, as a part of the Uruguay Round discussions, countries agreed to a significant overhaul of the MFA. First, the agreement was brought under the control of the WTO and renamed the Agreement on Textiles and Clothing (ATC). Second, countries decided to phase out the quotas completely over a ten-year transition period ending on January 1, 2005.

That transition to a quota-less industry did occur as scheduled; however, it is worth noting that many countries continue to maintain higher-than-average tariffs on textile and apparel products. Therefore, one still cannot say that free trade has been achieved.

Trade-Related Aspects of Intellectual Property Rights (TRIPS)

One major expansion of coverage of a trade liberalization agreement was the inclusion of intellectual property rights (IPR) into the discussion during the Uruguay Round. IPR covers the protections of written materials (copyrights), inventions (patents), and brand names and logos (trademarks). Most countries have established monopoly provisions for these types of creations in order to spur the creation of new writing and inventions and to protect the investments made in the establishment of trademarks. However, many of these protections have been unequally enforced around the world, resulting in a substantial amount of counterfeiting and pirating. The world is abound in fake CDs and DVDs, Gucci and Coach purses, and of course the international favorite, Rolex watches.

To harmonize the IPR protections around the world and to encourage enforcement of these provisions, countries created an IPR agreement called the Trade-Related Aspects of Intellectual Property Rights Agreement, or TRIPS. The TRIPS intends to both encourage trade and protect writers, inventors, and companies from the theft of their hard work and investments.

Other Agreements

What is listed and discussed above are just a few of the agreements negotiated during the Uruguay Round. In addition, any round of trade discussions provides an excellent forum for consideration of many other issues that are of particular interest to specific industries. Some of the others include the Agreement on Sanitary and Phytosanitary Measures, which provides guidelines for countries on food safety and plant and animal trade; an agreement on antidumping; the Agreement on Subsidies and Countervailing Measures; the Agreement on Trade-Related Investment Measures (TRIMS); the Agreement on Import-Licensing Procedures; the Agreement on Customs Valuation; the Preshipment Inspection Agreement; the Rules of Origin Agreement; and finally, several plurilateral agreements (meaning they don't cover everybody) concerning civilian aircraft, government procurement, and dairy products.

KEY TAKEAWAYS

- The Uruguay Round of the GATT resulted in numerous new trade-liberalizing agreements among member countries, including the General Agreement on Trade in Services (GATS), the Agreement on Agriculture, the Agreement on Textiles and Clothing (ATC), and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), among others.
- The GATS involved commitments to reduce regulations restricting international trade in services.
- The ATC involved commitments to eliminate the quota system established in the 1970s on textile and apparel products.
- The Agreement on Agriculture involved some modest commitments to reduce support for the agricultural industry.
- The TRIPS agreement involved commitments to standardize the treatment and enforcement of intellectual property rights.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The name of the U.S. legislation that prohibits foreign ships from transporting cargo between two U.S. ports.
 - b. The name used to describe services trade, such as language translations, provided by a foreign firm via the Internet.
 - c. The name used to describe services trade, such as banking, provided by a branch office located in the foreign country.
 - d. The name used to describe services trade, such as a hotel stay, provided to a foreigner traveling to the domestic country.
 - e. The name used to describe services trade, such as labor expertise, provided by foreign workers working in the domestic country.
 - f. The name of the Uruguay Round agreement liberalizing trade in services.
 - g. The name of the Uruguay Round agreement that superseded the Multi-Fiber Agreement (MFA).
 - h. The term used to describe the process of replacing import quotas with tariffs.
 - i. The name for a trade policy that sets a low tariff on a fixed quantity of imports and a high tariff on any imports over that quota.
 - j. The name of the Uruguay Round agreement on intellectual property rights.
 - k. The name of the Uruguay Round agreement on agriculture.

1.7 The World Trade Organization

LEARNING OBJECTIVE

1. Learn the basic intent of the World Trade Organization and its primary activities.

In order to monitor and sustain the complete set of Uruguay Round agreements, the member countries established a new body called the World Trade Organization (WTO). The WTO is a relatively small organization based in Geneva, Switzerland. It has a director-general, currently Pascal Lamy (as of January 2010), and a small staff of economists, lawyers, and others. The goal of the WTO is the same goal as its predecessor, the General Agreement on Tariffs and Trade (GATT): namely, to promote trade liberalization and thereby to foster growth and economic development.

Sometimes the WTO is described as an international organization governing international trade. However, this description can be misleading. The WTO does not make trade rules. The only makers of rules are national governments. In this sense, then, the WTO does not govern anybody. A better way to think of the WTO is as a club of member nations. The club's purpose is to monitor each member country's trade policies with respect to the trade agreements that were made in the Uruguay Round. The WTO agreements include thousands of promises for every country, all intending to reduce barriers to trade relative to what the barriers were before the Uruguay Round. The WTO does not represent free trade. At best, the agreements can be described as freer trade.

Besides monitoring each member country's trade policies, which the WTO fulfills by conducting periodic trade policy reviews of the member countries, the WTO club was also created to deal with disputes. This is surely the most important "power" of the WTO.

The Dispute Settlement Process

Disputes are handled by the Dispute Settlement Body (DSB). The DSB works like a committee that meets regularly to discuss any issues countries may have with respect to each other's trade policies. The DSB is comprised of one representative from each member country. When they meet, countries have the right to object to the trade policies of another country. However, they cannot object to anything or

everything; instead, a country can only object to an unfulfilled promise with respect to one or more of the WTO agreements.

When the Uruguay Round was finalized, each member country went back to its own legislature and changed its trade policies and rules to conform to its new commitments. Sometimes inadvertently and sometimes purposely, some countries do not implement their commitments fully. Or sometimes a country believes that it has fulfilled its commitment, but its trading partner believes otherwise. Or new legislation may violate one of the country's previous commitments. In these cases, a member country (the complainant) is allowed to register a dispute with the DSB against another member country (the defendant). Resolution of a dispute follows these steps:

1. *Consultations.* The DSB first demands that the appropriate government representatives from the complainant country and the defendant country meet to discuss the dispute. They must do this within a strict timetable (less than sixty days) and hopefully will be able to resolve the dispute without external intervention.
2. *Panel formation.* If the countries return to the DSB at a later session and report that the consultations failed, then the complainant may ask the DSB to form a panel. A panel consists of three to five independent trade law experts who are hired expressly to make a judgment about the particular dispute. The DSB chooses the panelists in consultation with the disputing countries, or the panelists are chosen by the director-general if the countries cannot agree. The panel is generally given about six months to decide whether the defendant violated some of its promises, whereupon it reports its decision to the DSB. Since a panel report can only be rejected by consensus, no country has veto power over DSB adoption of a report. Thus all panel reports become official decisions. But the process doesn't yet end.
3. *Appeals.* Either country can appeal the decision given in the panel report. A request or appeal sends the issue to an appellate board comprised of three judges drawn from a set of seven, each of whom has a four-year term. As in the U.S. court system, appellate arguments must be based on points of law relating to legal interpretations but cannot consider new evidence or retry the case. As with the original panel reports, appellate decisions are almost automatically adopted by the DSB.
4. *Resolution.* If the appellate board concurs with a panel decision that a defendant country has violated some of its WTO agreement commitments, there are two paths to resolution:

- a. *Compliance.* In the preferred outcome, the defendant country complies with the ruling against it and changes its laws as needed to conform. Sometimes compliance may take time because of delays in a legislative process, so normally the defendant will be given time to rectify the situation. In the process, the country will be expected to report its progress regularly to the DSB.
- b. *Suspension of concessions.* Sometimes a country refuses to comply with a ruling or it takes longer than the complainant is willing to wait. In this case, the complainant country is allowed by the DSB to suspend some of its previous concessions toward the defendant country. It works like this: Since it has been shown that the defendant has *not* lived up to all of its previous promises, the complainant is now allowed to rescind some of its own trade-liberalizing promises, but only toward the defendant country. To be fair, the rescission must have an effect on the defendant that is approximately equal in value to the cost imposed by the defendant's violations.

Dispute Settlement History

Since the WTO began in 1995 there have been over four hundred disputes brought to the DSB. A complete listing can be found at the WTO Web site here (http://www.wto.org/english/tratop_e/dispu_e/dispu_status_e.htm). A large number countries have been complainants and defendants although the two countries most often on one side or the other are the United States and the EU. Some of the most well-known disputes have involved bananas, steel, hormone-treated beef, and commercial aircraft. Lesser-known cases have involved narrow product groups such as Circular Welded Carbon Quality Line Pipe, Canned Tuna with Soybean Oil, Combed Cotton Yarn, and Retreaded Tires.

Many cases have been raised once, sent to consultations, and then never raised again. In some cases, consultations are sufficient to settle the dispute. Many other cases proceed to panel formation, appeals, and resolution. In many cases, defendants lose and eventually change their laws to comply with the WTO decision. In other cases, defendants lose and because of their refusal to comply, or their procrastination in complying, complainants suspend concessions. In a few cases, countries have refused to comply and faced no consequences. Occasionally, a defendant wins its case against a complainant.

Overall, the WTO dispute process has worked reasonably well. The cases brought, because they are often targeted to narrow industries, do not affect a huge amount of international trade. Nonetheless the existence of a forum in which to register

disputes and a mechanism for resolving them (one that includes some penalties for violations) has had a notable effect of reducing the risk of international trade.

Traders know better what to expect from their trading partners because their partners have committed themselves to particular trade policies and to a resolution mechanism in the event of noncompliance. In a sense, then, it is true that the WTO agreements restrict the freedom of a country to set whatever trade policy it deems appropriate for the moment. That loss of sovereignty, though, is designed to prevent countries from choosing more destructive protectionist policies—policies that are very seductive to voters, especially in an economic crisis. If successful, the WTO could prevent a reoccurrence of Smoot-Hawley and its aftermath both now and in the future.

KEY TAKEAWAYS

- The WTO's main purpose is to monitor the trade liberalization agreements reached by GATT member countries in the Uruguay Round.
- The most important “power” of the WTO is its ability to adjudicate disputes between member countries regarding compliance with the Agreements.
- Dispute resolution is conducted by the Dispute Settlement Body (DSB), which includes one representative from each WTO government.
- The four main steps to a WTO dispute case are (1) consultations, (2) panel formation, (3) appeals, and (4) resolution.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The name of the GATT round that created the WTO in 1995.
 - b. The name of the current director general of the WTO.
 - c. The term used to describe the process of rescinding one’s trade liberalization promises at the end of a WTO dispute.
 - d. The name of the WTO body that handles disagreements related to WTO commitments.
 - e. Countries must engage in these immediately after a dispute is raised at the WTO.
 - f. This official chooses dispute panel members if the complainant and defendant countries cannot agree.
 - g. The length of time served by a WTO appellate judge.
 - h. What a country is expected to do after losing a WTO dispute case.
 - i. The city in which WTO headquarters are located.
 - j. The approximate number of dispute cases filed at the WTO since its inception in 1995.

1.8 Appendix A: Selected U.S. Tariffs—2009

Table 1.2 "Special Tariff Classifications in the United States" contains a selection of the U.S. tariff rates specified in the 2009 U.S. Harmonized Tariff Schedule (HTS). The complete U.S. HTS is available at the U.S. International Trade Commission Web site (<http://www.usitc.gov>).

Table 1.2 Special Tariff Classifications in the United States

Symbol	Description
A, A*, A+	Generalized System of Preferences (GSP)
AU	U.S.-Australia free trade area (FTA)
B	Automotive Products Trade Act
BH	U.S.-Bahrain FTA
C	Agreement on Civil Aircraft
CA, MX	North American Free Trade Agreement (NAFTA): Canada and Mexico
CL	U.S.-Chile FTA
D	African Growth and Opportunity Act (AGOA)
E	Caribbean Basin Economic Recovery Act
IL	U.S.-Israel FTA
J, J*, J+	Andean Trade Preference Act
JO	U.S.-Jordan FTA
K	Agreement on Pharmaceuticals
P, P+	CAFTA-DR FTA
PE	U.S.-Peru FTA
MA	U.S.-Morocco FTA
OM	U.S.-Oman FTA
R	U.S.-Caribbean Trade Partnership Act
SG	U.S.-Singapore FTA

The tariff schedule in **Table 1.3 "Selected Tariffs in the United States, 2009"** displays four columns. The first column gives a brief description of the product. The second

column shows the product classification number. The first two numbers refer to the chapter, the most general product specification. For example, 08 refers to chapter 8, “Edible fruit and nuts; peel of citrus fruit or melons.” The product classification becomes more specific for each digit to the right. Thus 0805 refers more specifically to “Citrus fruit, fresh or dried.” The code 0805 40 refers to “Grapefruit,” and 0805 40 40 refers to “Grapefruit entering between August 1 and September 30.” This classification system is harmonized among about two hundred countries up to the first six digits and is overseen by the World Customs Organization.

The third column displays the “General Rate of Duty” for that particular product. This is the tariff that the United States applies to all countries with most-favored nation (MFN) status, or as it is now referred to in the United States, “normal trade relations” (NTR). The status was renamed NTR to provide a more accurate description of the term. One provision in the U.S. GATT/WTO agreements is that the United States promises to provide every WTO member country with MFN status. As a matter of policy, the United States also typically grants most non-WTO countries the same status. For example, as of 2009, Russia was not a member of the WTO, but the United States applied its NTR tariff rates to Russian imports.

The final column lists special rates of duty that apply to select countries under special circumstances. For each product, you will see a tariff rate followed by a list of symbols in parentheses. The symbols indicate the trade act or free trade agreement that provides special tariff treatment to those countries. A complete list of these is shown in [Table 1.2 "Special Tariff Classifications in the United States"](#). Symbols that include a “+” or “*” generally refer to special exceptions that apply for some countries with that product.

In the standard U.S. tariff schedule, there is one additional column labeled “2.” This is the U.S. non-MFN tariff, meaning essentially the nonspecial tariffs. Many of these tariff rates, especially for product categories that have been around for a long time, are holdovers from the Smoot-Hawley tariffs set in the Tariff Act of 1930. They are significantly higher than the standard MFN tariffs in column 1 but apply to only two countries: Cuba and North Korea.

Table 1.3 Selected Tariffs in the United States, 2009

Description	HTS Code	MFN/NTR Tariff	Special Tariff
Cauliflower, broccoli	0704.10.20	2.5% (June 5–Oct. 25)	Free (A,AU,BH,CA,CL,E,IL,J,JO,MA,MX,OM,P,PE,SG)

Description	HTS Code	MFN/NTR Tariff	Special Tariff
	0704.10.40	10% (Other, not reduced in size)	Free (A,AU,BH,CA,CL,E,IL,J,JO,MA,MX,OM,P,PE,SG)
	0704.10.60	14% (Cut or sliced)	Free (A,BH,CA,CL,E,IL,J,JO,MA,MX,OM,P,PE) 7% (AU) 3.5% (SG)
Grapefruit, incl. pomelos	0805.40.40	1.9¢/kg (Aug.–Sept.)	Free (AU,BH,CA,D,E,IL,J,JO,MA,MX,OM,P,PE,SG)
	0805.40.60	1.5¢/kg (Oct.)	Free (CA, CL, D, E,IL,J,JO,MX,P,PE, SG) 1¢/kg (AU) 0.9¢/kg (BH) 1.1¢/kg (MA) 1.2¢/kg (OM)
	0805.40.80	2.5¢/kg (Nov.–July)	Free (CA, D, E, IL, J, JO, MX, P, PE) 1.8¢/kg (AU,MA) 1.5¢/kg (BH) 1¢/kg (CL,SG) 2.2¢/kg (OM)

Description	HTS Code	MFN/NTR Tariff	Special Tariff
Grapes, fresh	0806.10.20	\$1.13/m ³ (Feb. 15–Mar. 31)	Free (A+, AU, BH, CA, CL, D, E, IL, J, JO, MA, MX, OM, P, PE, SG)
	0806.10.40	Free (Apr. 1–June 30)	
	0806.10.60	\$1.80/m ³ (any other time)	Free (A+, AU, BH, CA, CL, D, E, IL, J, JO, MA, MX, OM, P, PE, SG)
Ceramic tableware; cups valued over \$5.25 per dozen; saucers valued over \$3 per dozen; soups, oatmeals, and cereals valued over \$6 per dozen; plates not over 22.9 cm in maximum diameter and valued over \$6 per dozen; plates over 22.9 but not over 27.9 cm in maximum diameter and valued over \$8.50 per dozen; platters or chop dishes valued over \$35 per dozen; sugars valued over \$21 per dozen; creamers	6912.00.45	4.5%	Free (A+, AU, CA, CL, D, E, IL, J, JO, MX, P, PE, SG) 2.7% (BH) 2.4% (MA) 4% (OM)

Description	HTS Code	MFN/NTR Tariff	Special Tariff
valued over \$15 per dozen; and beverage servers valued over \$42 per dozen			
Motor cars principally designed for the transport of persons, of all cylinder capacities	8703.2x.00	2.5%	Free (A+,AU,B,BH,CA,CL,D,E,IL,J,JO,MA,MX,OM,P,PE,SG)
Motor vehicles for the transport of goods (i.e., trucks), gross vehicle weight exceeding 5 metric tons but less than 20 metric tons	8704.22.50	25%	Free (A+,AU,B,BH,CA,CL,D,E,IL,J,MA,MX,OM,P,PE) 2.5% (JO) 10% (SG)
Bicycles having both wheels not exceeding 63.5 cm in diameter	8712.00.15	11%	Free (A+,AU,BH,CA,CL,D,E,IL,J,JO,MA,MX,OM,P,PE) 1.3% (SG)
Cane sugar	1701.11.05	1.4606¢/kg less 0.020668¢/kg for each degree under 100 degrees but not less than 0.943854¢/kg	Free (A*,AU,BH,CA,CL,E*,IL,J,JO,MA,MX,OM,P,PE,SG)

Description	HTS Code	MFN/NTR Tariff	Special Tariff
Sports footwear: tennis shoes, basketball shoes, gym shoes, training shoes and the like: having uppers of which over 50% of the external surface area is leather	6404.11.20	10.5%	Free (AU,BH,CA,CL,D,E,IL,J+,JO,MA,MX,OM,P,PE,R) 1.3% (SG)
Golf clubs	9506.31.00	4.4%	Free (A,AU,BH,CA,CL,E,IL,J,JO,MA,MX,OM,P,PE,SG)
Wristwatches	9101.11.40	51¢ each + 6.25% on case and strap + 5.3% on battery	Free (AU,BH,CA,CL,D,E,IL,J,J+,JO,MA,MX,OM,P,PE,R,SG)
Fax machines	8517.21.00	Free	
Coffee, caffeinated	0901.21.00	Free	
Tea, green tea, flavored	0902.10.10	6.4%	Free (A,AU,BH,CA,CL,E,IL,J,JO,MA,MX,OM,P,PE,SG)

The products presented in [Table 1.3 "Selected Tariffs in the United States, 2009"](#) were selected to demonstrate several noteworthy features of U.S. trade policy. The WTO reports in the 2006 U.S. trade policy review that most goods enter the United States either duty free or with very low tariffs. Coffee and fax machines are two goods, shown above, representative of the many goods that enter duty free. The average MFN tariff in the United States in 2002 was about 5 percent, although for agricultural goods the rate was almost twice as high. About 7 percent of U.S. tariffs exceed 15 percent; these are mostly sensitive products such as peanuts, dairy, footwear, textiles, and clothing. The trade-weighted average tariff in the United States was only about 1.5 percent in 2003.

One interesting feature of the tariff schedule is the degree of specificity of the products in the HTS schedule. Besides product type, categories are divided according to weight, size, or the time of year. Note especially the description of ceramic tableware and bicycles.

Tariffs vary according to time of entry, as with cauliflower, grapefruit, and grapes. This reflects the harvest season for those products in the United States. When the tariff is low, that product is out of season in the United States. Higher tariffs are in place when U.S. output in the product rises.

Notice the tariffs on cauliflower and broccoli. They are lower if the vegetables are unprocessed. If the product is cut or sliced before arriving in the United States, the tariff rises to 14 percent. This reflects a case of tariff escalation. Tariff escalation means charging a higher tariff the greater the degree of processing for a product. This is a common practice among many developed countries and serves to protect domestic processing industries. Developing countries complain that these practices impede their development by preventing them from competing in more advanced industries. Consequently, tariff escalation is a common topic of discussion during trade liberalization talks.

Tariff rates also vary with different components of the same product, as with watches. Note also that watches have both specific tariffs and ad valorem tariffs applied.

Notice that the tariff on cars in the United States is 2.5 percent, but the tariff on truck imports is ten times that rate at 25 percent. The truck tariff dates back to 1963 and is sometimes referred to as the “chicken tax.” It was implemented primarily to affect Volkswagen in retaliation for West Germany’s high tariff on chicken imports from the United States. Today, Canada and Mexico are exempt from the tariff due to NAFTA, and Australia will also be exempt with the new U.S.-Australia FTA. The truck tax is set to be a contentious issue in current U.S.-Thailand FTA discussions.

The tariff rates themselves are typically set to several significant digits. One has to wonder why the United States charges 4.4 percent on golf clubs rather than an even 4 percent or 5 percent. Much worse is the tariff rate on cane sugar with six significant digits.

The special tariff rates are often labeled “free,” meaning these goods enter duty-free from that group of countries. Note that Chile and Singapore sometimes have tariff rates in between the MFN rate and zero. This reflects the FTA’s phase in the

process. Most FTAs include a five- to fifteen-year phase-in period during which time tariffs are reduced annually toward zero.

One thing to think about while reviewing this tariff schedule is the administrative cost of monitoring and taxing imported goods. Not only does the customs service incur costs to properly categorize and measure goods entering the country, but foreign firms themselves must be attuned to the intricacies of the tariff schedule of all the countries to which they export. All of this requires the attention and time of employees of the firms and represents a cost of doing business. These administrative costs are rarely included in the evaluation of trade policies.

An administratively cheaper alternative would be to charge a fixed ad valorem tariff on all goods that enter, much like a local sales tax. However, for political reasons, it would be almost impossible to switch to this much simpler alternative.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?” [Note: the following exercises are meant to provide practice in reading and interpreting the U.S. tariff schedule.]
 - a. The 2009 MFN tariff rate on imported broccoli that has been processed by cutting or slicing before shipping.
 - b. The allowable diameter range for ceramic plates valued over \$8.50 under HTS code 6912.00.45.
 - c. The 2009 U.S. tariff on truck imports from Singapore.
 - d. The 2009 MFN tariff on cauliflower that entered the U.S. in November.
 - e. The 2009 U.S. tariff on golf clubs from Israel.

1.9 Appendix B: Bound versus Applied Tariffs

The WTO agreement includes commitments by countries to bind their tariff rates at an agreed-upon maximum rate for each import product category. The maximum tariff in a product category is called the *bound tariff rate*. The bound tariff rates differ across products and across countries: some countries agree to higher maximums; others agree to lower maximums. In general, less-developed countries have higher bound tariff rates than developed countries, reflecting their perception that they need greater protection from competition against the more highly developed industries in the developed markets.

However, some countries, especially those with higher bound tariffs, decide to set their actual tariffs at lower levels than their bound rates. The actual tariff rate is called the *applied tariff rate*. [Table 1.4 "Bound versus Applied Average Tariffs"](#) lists the average applied tariff rates compared to average bound tariffs for a selected set of WTO member countries. The averages are calculated as a simple average: namely, the ad valorem tariff rates (bound or applied) are added together and divided by the total number of tariff categories. These are not trade-weighted average tariffs. Also, when specific tariffs are assessed for a product, they are excluded from the calculations. (Note that specific tariffs are set as a dollar charge per unit of imports.) Also listed is the percentage of six-digit tariff lines that have a tariff binding. For products that have no tariff binding, the country is free to set whatever tariff it wishes. The countries are ordered from the highest to the lowest gross domestic product (GDP) per person.

Table 1.4 Bound versus Applied Average Tariffs

Country	Applied Rate (%)	Bound Rate (%)	% Bound
United States	3.6	3.6	100.0
Canada	3.6	5.1	99.7
EC	4.3	4.1	100.0
Japan	3.1	2.9	99.6
South Korea	11.3	16.0	94.7
Mexico	12.5	34.9	100.0
Chile	6.0 (uniform)	25.1	100.0
Argentina	11.2	32.0	100.0
Brazil	13.6	31.4	100.0

Country	Applied Rate (%)	Bound Rate (%)	% Bound
Thailand	9.1	25.7	74.7
China	9.95	10.0	100.0
Egypt	17.0	36.8	99.3
Philippines	6.3	25.6	66.8
India	15.0	49.7	73.8
Kenya	12.7	95.7	14.6
Ghana	13.1	92.5	14.3

Table 1.4 "Bound versus Applied Average Tariffs" reveals the following things worth noting:

1. More-developed countries tend to apply lower average tariffs than less-developed countries (LDCs).
2. Average bound tariff rates are higher for less-developed countries. This means that the WTO agreement has not forced LDCs to open their economies to the same degree as developed countries.
3. The less developed a country, the fewer tariff categories that are bound. For the most developed economies, 100 percent of the tariff lines are bound, but for Ghana and Kenya, only 14 percent are bound. This also means that the WTO agreement has not forced LDCs to open their economies to the same degree as developed countries.
4. For LDCs, applied tariffs are set much lower on average than the bound rates. These countries have the flexibility to raise their tariffs without violating their WTO commitments.
5. China has lower tariffs and greater bindings than countries of similar wealth.
6. Since the most developed economies have applied rates equal to bound rates, they cannot raise tariffs without violating their WTO commitments. WTO-sanctioned trade remedy actions can be used instead, however.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term for the maximum tariff rate a country agrees to assess on imports from other WTO member countries.
 - b. The term for the actual tariff rate a country assesses on imports from other WTO member countries.
 - c. Between *developed* or *less developed countries*, these tend to have much higher bound tariff rates.
 - d. The percentage of tariff lines on which the Philippines has agreed to set maximum tariffs in the WTO.
 - e. The average WTO-bound tariff rate in Ghana.
 - f. One country that has agreed to much lower bound tariffs than other countries of comparable income and wealth in the WTO.

Chapter 2

The Ricardian Theory of Comparative Advantage

This chapter presents the first formal model of international trade: the Ricardian model. It is one of the simplest models, and still, by introducing the principle of comparative advantage, it offers some of the most compelling reasons supporting international trade. Readers will learn some of the surprising outcomes of the Ricardian model; for example, less productive nations can benefit from free trade with their more productive neighbors, and very low-wage countries are unlikely to be able to use their production cost advantage in many circumstances. Readers will also learn why so many people, even those who have studied the Ricardian theory, consistently get the results wrong.

In other words, the Ricardian model is both one of the most misunderstood and one of the most compelling models of international trade.

2.1 The Reasons for Trade

LEARNING OBJECTIVES

1. Learn the five reasons why trade between countries may occur.
2. Recognize that separate models of trade incorporate different motivations for trade.

The first theory section of this course develops models that provide different explanations or reasons why trade takes place between countries. The five basic reasons why trade may take place are summarized below. The purpose of each model is to establish a basis for trade and then to use that model to identify the expected effects of trade on prices, profits, incomes, and individual welfare.

Reason for Trade #1: Differences in Technology

Advantageous trade can occur between countries if the countries differ in their technological abilities to produce goods and services. Technology refers to the techniques used to turn resources (labor, capital, land) into outputs (goods and services). The basis for trade in the *Ricardian model of comparative advantage* in [Chapter 2 "The Ricardian Theory of Comparative Advantage"](#) is differences in technology.

Reason for Trade #2: Differences in Resource Endowments

Advantageous trade can occur between countries if the countries differ in their endowments of resources. Resource endowments refer to the skills and abilities of a country's workforce, the natural resources available within its borders (minerals, farmland, etc.), and the sophistication of its capital stock (machinery, infrastructure, communications systems). The basis for trade in both the pure exchange model in [Chapter 3 "The Pure Exchange Model of Trade"](#) and the Heckscher-Ohlin model in [Chapter 5 "The Heckscher-Ohlin \(Factor Proportions\) Model"](#) is differences in resource endowments.

Reason for Trade #3: Differences in Demand

Advantageous trade can occur between countries if demands or preferences differ between countries. Individuals in different countries may have different preferences or demands for various products. For example, the Chinese are likely to

demand more rice than Americans, even if consumers face the same price. Canadians may demand more beer, the Dutch more wooden shoes, and the Japanese more fish than Americans would, even if they all faced the same prices. There is no formal trade model with demand differences, although the monopolistic competition model in [Chapter 6 "Economies of Scale and International Trade"](#) does include a demand for variety that can be based on differences in tastes between consumers.

Reason for Trade #4: Existence of Economies of Scale in Production

The existence of economies of scale in production is sufficient to generate advantageous trade between two countries. Economies of scale refer to a production process in which production costs fall as the scale of production rises. This feature of production is also known as “increasing returns to scale.” Two models of trade incorporating economies of scale are presented in [Chapter 6 "Economies of Scale and International Trade"](#).

Reason for Trade #5: Existence of Government Policies

Government tax and subsidy programs alter the prices charged for goods and services. These changes can be sufficient to generate advantages in production of certain products. In these circumstances, advantageous trade may arise solely due to differences in government policies across countries. [Chapter 8 "Domestic Policies and International Trade"](#), [Section 8.3 "Production Subsidies as a Reason for Trade"](#) and [Chapter 8 "Domestic Policies and International Trade"](#), [Section 8.6 "Consumption Taxes as a Reason for Trade"](#) provide several examples in which domestic tax or subsidy policies can induce international trade.

Summary

There are very few models of trade that include all five reasons for trade simultaneously. The reason is that such a model is too complicated to work with. Economists simplify the world by choosing a model that generally contains just one reason. This does not mean that economists believe that one reason, or one model, is sufficient to explain all outcomes. Instead, one must try to understand the world by looking at what a collection of different models tells us about the same phenomenon.

For example, the Ricardian model of trade, which incorporates differences in technologies between countries, concludes that everyone benefits from trade, whereas the Heckscher-Ohlin model, which incorporates endowment differences,

concludes that there will be winners and losers from trade. Change the basis for trade and you may change the outcomes from trade.

In the real world, trade takes place because of a combination of all these different reasons. Each single model provides only a glimpse of some of the effects that might arise. Consequently, we should expect that a combination of the different outcomes that are presented in different models is the true characterization of the real world. Unfortunately, because of this, understanding the complexities of the real world is still more of an art than a science.

KEY TAKEAWAYS

- The five main reasons international trade takes place are differences in technology, differences in resource endowments, differences in demand, the presence of economies of scale, and the presence of government policies.
- Each model of trade generally includes just one motivation for trade.

EXERCISES

1. List the five reasons why international trade takes place.
2. Identify which model incorporates
 - a. differences in technology,
 - b. presence of economies of scale,
 - c. differences in demand,
 - d. differences in endowments.

2.2 The Theory of Comparative Advantage: Overview

LEARNING OBJECTIVES

1. Learn how a rearrangement of production on the basis of comparative advantage, coupled with international trade, can lead to an improvement in the well-being of individuals in all countries.
2. Learn the major historical figures who first described the effects of international trade: Adam Smith, David Ricardo, and Robert Torrens.

Historical Overview

The theory of **comparative advantage**¹ is perhaps the most important concept in international trade theory. It is also one of the most commonly misunderstood principles. There is a popular story told among economists that once when an economics skeptic asked Paul Samuelson (a Nobel laureate in economics) to provide a meaningful and nontrivial result from the economics discipline, Samuelson quickly responded, “comparative advantage.”

The sources of the misunderstandings are easy to identify. First, the principle of comparative advantage is clearly counterintuitive. Many results from the formal model are contrary to simple logic. Second, it is easy to confuse the theory with another notion about advantageous trade, known in trade theory as the theory of absolute advantage. The logic behind absolute advantage is quite intuitive. This confusion between these two concepts leads many people to think that they understand comparative advantage when in fact what they understand is absolute advantage. Finally, the theory of comparative advantage is all too often presented only in its mathematical form. Numerical examples or diagrammatic representations are extremely useful in demonstrating the basic results and the deeper implications of the theory. However, it is also easy to see the results mathematically without ever understanding the basic intuition of the theory.

1. A country has a comparative advantage when it can produce a good at a lower opportunity cost than another country; alternatively, when the relative productivities between goods compared with another country are the highest.

The early logic that free trade could be advantageous for countries was based on the concept of absolute advantages in production. Adam Smith wrote in *The Wealth of Nations*, “If a foreign country can supply us with a commodity cheaper than we ourselves can make it, better buy it of them with some part of the produce of our own industry, employed in a way in which we have some advantage” (Book IV, Section ii, 12). For more information, see Rod Hay, “Adam Smith,” McMaster

University Archive for the History of Economic Thought,
<http://socserv.mcmaster.ca/econ/ugcm/3ll3.smith/wealth/index.html>.

The idea here is simple and intuitive. If our country can produce some set of goods at a lower cost than a foreign country and if the foreign country can produce some other set of goods at a lower cost than we can produce them, then clearly it would be best for us to trade our relatively cheaper goods for their relatively cheaper goods. In this way, both countries may gain from trade.

The original idea of comparative advantage dates to the early part of the nineteenth century. For a more complete history of these ideas, see Douglas A. Irwin, *Against the Tide: An Intellectual History of Free Trade* (Princeton, NJ: Princeton University Press, 1996). Although the model describing the theory is commonly referred to as the "Ricardian model," the original description of the idea (see [Chapter 2 "The Ricardian Theory of Comparative Advantage"](#), [Section 2.12 "Appendix: Robert Torrens on Comparative Advantage"](#)) can be found in the 1815 *Essay on the External Corn Trade*. See Robert Torrens, *Essay on the External Corn Trade* (London: J. Hatchard, 1815) by Robert Torrens. David Ricardo formalized the idea using a compelling yet simple numerical example in his 1817 book *On the Principles of Political Economy and Taxation*. See David Ricardo, *On the Principles of Political Economy and Taxation*, McMaster University Archive for the History of Economic Thought, <http://socserv2.socsci.mcmaster.ca/~econ/ugcm/3ll3/ricardo/prin/index.html>. The idea appeared again in James Mill's 1821 *Elements of Political Economy*. See James Mill, *Elements of Political Economy* (London: Baldwin, Cradock & Joy, 1821). Finally, the concept became a key feature of international political economy upon the 1848 publication of *Principles of Political Economy* by John Stuart Mill. See John Stuart Mill, *Principles of Political Economy*, McMaster University Archive for the History of Economic Thought, <http://socserv2.socsci.mcmaster.ca/~econ/ugcm/3ll3/mill/index.html>.

Ricardo's Numerical Example

Because the idea of comparative advantage is not immediately intuitive, the best way of presenting it seems to be with an explicit numerical example as provided by Ricardo. Indeed, some variation of Ricardo's example lives on in most international trade textbooks today.

In his example, Ricardo imagined two countries, England and Portugal, producing two goods, cloth and wine, using labor as the sole input in production. He assumed that the productivity of labor (i.e., the quantity of output produced per worker) varied between industries and across countries. However, instead of assuming, as Adam Smith did, that England is more productive in producing one good and

Portugal is more productive in the other, Ricardo assumed that Portugal was more productive in both goods. Based on Smith's intuition, then, it would seem that trade could not be advantageous, at least for England.

However, Ricardo demonstrated numerically that if England specialized in producing one of the two goods and if Portugal produced the other, then total world output of both goods could rise! If an appropriate **terms of trade**² (i.e., amount of one good traded for another) were then chosen, both countries could end up with more of both goods after specialization and free trade than they each had before trade. This means that England may nevertheless benefit from free trade even though it is assumed to be technologically inferior to Portugal in the production of everything.

As it turned out, specialization in *any* good would not suffice to guarantee the improvement in world output. Only one of the goods would work. Ricardo showed that the specialization good in each country should be that good in which the country had a comparative advantage in production. To identify a country's comparative advantage good requires a comparison of production costs across countries. However, one does not compare the monetary costs of production or even the resource costs (labor needed per unit of output) of production. Instead, one must compare the *opportunity costs* of producing goods across countries.

A country is said to have a comparative advantage in the production of a good (say, cloth) if it can produce it at a lower opportunity cost than another country. The opportunity cost of cloth production is defined as the amount of wine that must be given up in order to produce one more unit of cloth. Thus England would have the comparative advantage in cloth production relative to Portugal if it must give up less wine to produce another unit of cloth than the amount of wine that Portugal would have to give up to produce another unit of cloth.

All in all, this condition is rather confusing. Suffice it to say that it is quite possible, indeed likely, that although England may be less productive in producing both goods relative to Portugal, it will nonetheless have a comparative advantage in the production of one of the two goods. Indeed, there is only one circumstance in which England would not have a comparative advantage in either good, and in this case Portugal also would not have a comparative advantage in either good. In other words, either each country has the comparative advantage in one of the two goods or neither country has a comparative advantage in anything.

2. The amount of one good traded per unit of another in a mutually voluntary exchange. Often expressed as a ratio of prices and measured as a ratio of units; for example, pounds of cheese per gallon of wine.

Another way to define comparative advantage is by comparing productivities across industries and countries. Suppose, as before, that Portugal is more productive than England in the production of both cloth and wine. If Portugal is twice as productive

in cloth production relative to England but three times as productive in wine, then Portugal's comparative advantage is in wine, the good in which its productivity advantage is greatest. Similarly, England's comparative advantage good is cloth, the good in which its productivity disadvantage is least. This implies that to benefit from specialization and free trade, Portugal should specialize in and trade the good that it is "most better" at producing, while England should specialize in and trade the good that it is "least worse" at producing.

Note that trade based on comparative advantage does not contradict Adam Smith's notion of advantageous trade based on absolute advantage. If, as in Smith's example, England were more productive in cloth production and Portugal were more productive in wine, then we would say that England has an absolute advantage in cloth production, while Portugal has an absolute advantage in wine. If we calculated comparative advantages, then England would also have the comparative advantage in cloth and Portugal would have the comparative advantage in wine. In this case, gains from trade could be realized if both countries specialized in their comparative and absolute advantage goods. Advantageous trade based on comparative advantage, then, covers a larger set of circumstances while still including the case of absolute advantage and hence is a more general theory.

The Ricardian Model: Assumptions and Results

The modern version of the Ricardian model and its results is typically presented by constructing and analyzing an economic model of an international economy. In its most simple form, the model assumes two countries producing two goods using labor as the only factor of production. Goods are assumed to be **homogeneous**³ (i.e., identical) across firms and countries. Labor is homogeneous within a country but heterogeneous (nonidentical) across countries. Goods can be transported costlessly between countries. Labor can be reallocated costlessly between industries within a country but cannot move between countries. Labor is always fully employed. Production technology differences exist across industries and across countries and are reflected in labor productivity parameters. The labor and goods markets are assumed to be perfectly competitive in both countries. Firms are assumed to maximize profit, while consumers (workers) are assumed to maximize utility.

The primary issue in the analysis of this model is what happens when each country moves from **autarky**⁴ (no trade) to free trade with the other country—in other words, what are the effects of trade? The main things we care about are trade's effects on the prices of the goods in each country, the production levels of the goods, employment levels in each industry, the pattern of trade (who exports and who imports what), consumption levels in each country, wages and incomes, and the welfare effects both nationally and individually.

3. Goods, or production factors, that are identical and thus perfectly substitutable in consumption, or production.

4. The situation in which a country does not trade with the rest of the world.

Using the model, one can show that in autarky each country will produce some of each good. Because of the technology differences, relative prices of the two goods will differ between countries. The price of each country's comparative advantage good will be lower than the price of the same good in the other country. If one country has an absolute advantage in the production of both goods (as assumed by Ricardo), then real wages of workers (i.e., the purchasing power of wages) in that country will be higher in both industries compared to wages in the other country. In other words, workers in the technologically advanced country would enjoy a higher standard of living than in the technologically inferior country. The reason for this is that wages are based on productivity; thus in the country that is more productive, workers get higher wages.

The next step in the analysis is to assume that trade between countries is suddenly liberalized and made free. The initial differences in relative prices of the goods between countries in autarky will stimulate trade between the countries. Since the differences in prices arise directly out of differences in technology between countries, it is the differences in technology that cause trade in the model. Profit-seeking firms in each country's comparative advantage industry would recognize that the price of their good is higher in the other country. Since transportation costs are zero, more profit can be made through export than with sales domestically. Thus each country would export the good in which it has a comparative advantage. Trade flows would increase until the price of each good is equal across countries. In the end, the price of each country's export good (its comparative advantage good) will rise and the price of its import good (its comparative disadvantage good) will fall.

The higher price received for each country's comparative advantage good would lead each country to specialize in that good. To accomplish this, labor would have to move from the comparative disadvantage industry into the comparative advantage industry. This means that one industry goes out of business in each country. However, because the model assumes full employment and costless mobility of labor, all these workers are immediately gainfully employed in the other industry.

One striking result here is that even when one country is technologically superior to the other in both industries, one of these industries would go out of business when opening to free trade. Thus technological superiority is not enough to guarantee continued production of a good in free trade. A country must have a comparative advantage in production of a good rather than an absolute advantage to guarantee continued production in free trade. From the perspective of a less-developed country, *the developed country's superior technology need not imply that less-developed country (LDC) industries cannot compete in international markets.*

Another striking result is that the technologically superior country's comparative advantage industry survives while the same industry disappears in the other country, even though the workers in the other country's industry have lower wages. In other words, low wages in another country in a particular industry is not sufficient information to determine which country's industry would perish under free trade. From the perspective of a developed country, *free trade may not result in a domestic industry's decline just because the foreign firms pay their workers lower wages.*

The movement to free trade generates an improvement in welfare in both countries individually and nationally. Specialization and trade will increase the set of consumption possibilities, compared with autarky, and will make possible an increase in consumption of both goods nationally. These aggregate gains are often described as improvements in production and consumption efficiency. Free trade raises aggregate world production efficiency because more of both goods are likely to be produced with the same number of workers. Free trade also improves aggregate consumption efficiency, which implies that consumers have a more pleasing set of choices and prices available to them.

Real wages (and incomes) of individual workers are also shown to rise in both countries. Thus every worker can consume more of both goods in free trade compared with autarky. In short, everybody benefits from free trade in both countries. In the Ricardian model, trade is truly a win-win situation.

Defending against Skeptics: The Intuition behind the Theory of Comparative Advantage

Many people who learn about the theory of comparative advantage quickly convince themselves that its ability to describe the real world is extremely limited, if not nonexistent. Although the results follow logically from the assumptions, the assumptions are easily assailed as unrealistic. For example, the model assumes only two countries producing two goods using just one factor of production. No capital or land or other resources are needed for production. The real world, on the other hand, consists of many countries producing many goods using many factors of production. In the model, each market is assumed to be perfectly competitive when in reality there are many industries in which firms have market power. Labor productivity is assumed to be fixed when in actuality it changes over time, perhaps based on past production levels. Full employment is assumed when clearly workers cannot immediately and costlessly move to other industries. Also, all workers are assumed to be identical. This means that when a worker is moved from one industry to another, he or she is immediately as productive as every other worker who was previously employed there. Finally, the model assumes that technology differences are the only differences that exist between the countries.

With so many unrealistic assumptions, it is difficult for some people to accept the conclusions of the model with any confidence, especially when so many of the results are counterintuitive. Indeed, one of the most difficult aspects of economic analysis is how to interpret the conclusions of models. Models are, by their nature, simplifications of the real world and thus all economic models contain unrealistic assumptions. Therefore, to dismiss the results of economic analysis on the basis of unrealistic assumptions means that one must dismiss all insights contained within the entire economics discipline. Surely, this is neither practical nor realistic. Economic models in general and the Ricardian model in particular do contain insights that most likely carry over to the more complex real world. The following story is meant to explain some of the insights within the theory of comparative advantage by placing the model into a more familiar setting.

A Gardening Story

Suppose it is early spring and it is time to prepare the family backyard garden for the first planting of the year. The father in the household sets aside one Sunday afternoon to do the job but hopes to complete the job as quickly as possible. Preparation of the garden requires the following tasks. First, the soil must be turned over and broken up using the rototiller. Then the soil must be raked and smoothed. Finally, seeds must be planted, or sowed.

This year, the father's seven-year-old son is anxious to help. The question at hand is whether the son should be allowed to help if one's only objective is to complete the task in the shortest amount of time possible.

At first thought, the father is reluctant to accept help. Clearly each task would take the father less time to complete than it would take the son. In other words, the father can perform each task more efficiently than the seven-year-old son. The father estimates that it will take him three hours to prepare the garden if he works alone, as shown in [Table 2.1 "Father's Task Times without Son"](#).

Table 2.1 Father's Task Times without Son

Task	Completion Time (Hours)
Rototilling	1.0
Raking	1.0
Planting	1.0
Total	3.0

On second thought, the father decides to let his son help according to the following procedure. First, the father begins the rototilling. Once he has completed half of the garden, the son begins raking the rototilled section while the father finishes rototilling the rest of the garden plot. After the father finishes rototilling, he begins planting seeds in the section the son has already raked. Suppose that the son rakes slower than the father plants and that the father completes the sowing process just as the son finishes raking. Note this implies that raking takes the son almost two hours compared to one hour for the father. However, because the son's work and the father's work are done simultaneously, it does not add to the total time for the project. Under this plan, the time needed to complete the tasks is shown in [Table 2.2 "Father's Task Times with Son"](#).

Table 2.2 Father's Task Times with Son

Task	Completion Time (Hours)
Rototilling	1.0
Raking and Planting	1.0
Total	2.0

Notice that the total time needed to prepare the garden has fallen from three hours to two hours. The garden is prepared in less time with the son's help than it could have been done independently by the father. In other words, it makes sense to employ the son in (garden) production even though the son is less efficient than the dad in every one of the three required tasks. Overall efficiency is enhanced when both resources (the father and son) are fully employed.

This arrangement also clearly benefits both the father and son. The father completes the task in less time and thus winds up with some additional leisure time that the father and son can enjoy together. The son also benefits because he has contributed his skills to a productive activity and will enjoy a sense of accomplishment. Thus both parties benefit from the arrangement.

However, it is important to allocate the tasks correctly between the father and the son. Suppose the father allowed his son to do the rototilling instead. In this case, the time needed for each task might look as it does in [Table 2.3 "Task Times with Incorrect Specialization"](#).

Table 2.3 Task Times with Incorrect Specialization

Task	Completion Time (Hours)
Rototilling	4.0
Raking	1.0
Planting	1.0
Total	6.0

The time needed for rototilling has now jumped to four hours because we have included the time spent traveling to and from the hospital and the time spent in the emergency room! Once the father and son return, the father must complete the remaining tasks on his own. Overall efficiency declines in this case compared with the father acting alone.

This highlights the importance of specializing in production of the task in which you have a comparative advantage. Even though the father can complete all three tasks quicker than his son, his relative advantage in rototilling greatly exceeds his advantage in raking and planting. One might say that the father is “most better” at rototilling, while he is “least better” at raking and planting. On the other hand, the son is “least worse” at raking and planting but “most worse” at rototilling. Finally, because of the sequential nature of the tasks, the son can remain fully employed only if he works on the middle task, namely, raking.

Interpreting the Theory of Comparative Advantage

The garden story offers an intuitive explanation for the theory of comparative advantage and also provides a useful way of interpreting the model results. The usual way of stating the Ricardian model results is to say that countries *will* specialize in their comparative advantage good and trade it to the other country such that everyone in both countries benefits. Stated this way, it is easy to imagine how it would not hold true in the complex real world.

A better way to state the results is as follows. The Ricardian model shows that if we want to maximize total output in the world, then we should

1. fully employ all resources worldwide,
2. allocate those resources within countries to each country’s comparative advantage industries,
3. allow the countries to trade freely thereafter.

In this way, we might raise the well-being of all individuals despite differences in relative productivities. In this description, we do not predict that a result will carry over to the complex real world. Instead, we carry the logic of comparative advantage to the real world and ask how things would have to look to achieve a certain result (maximum output and benefits). In the end, we should not say that the model of comparative advantage tells us anything about what *will* happen when two countries begin to trade; instead, we should say that the theory tells us some things that *can* happen.

KEY TAKEAWAYS

- Trade based on comparative advantage can make everyone in both countries better off after trade.
- Superior technology in developed countries need not imply that industries in less-developed countries cannot compete in international markets.
- Firms in developed countries can sometimes compete in international markets even when foreign firms pay their workers much lower wages.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe workers who have the same productivity in multiple industries.
 - b. The term used to describe a product when it is identical across multiple firms.
 - c. The term used to describe a product, like wine, that is produced by different firms, each with slightly different characteristics.
 - d. The assumption made about labor employment in the Ricardian model.
 - e. The term used to describe the amount of goods that can be produced using all the available world resources.
2. What three things must be achieved to maximize world output?
3. In the gardening story, if the son can do the rototilling in four hours, the raking in two hours, and the planting in three hours, which activity is the son “least worse” in producing compared with his father?

2.3 Ricardian Model Assumptions

LEARNING OBJECTIVE

1. Learn the structure and assumptions that describe the Ricardian model of comparative advantage.

The Ricardian model shows the possibility that an industry in a developed country could compete against an industry in a less-developed country (LDC) even though the LDC industry pays its workers much lower wages.

The modern version of the Ricardian model assumes that there are two countries producing two goods using one factor of production, usually labor. The model is a general equilibrium model in which all markets (i.e., goods and factors) are perfectly competitive. The goods produced are assumed to be homogeneous across countries and firms within an industry. Goods can be costlessly shipped between countries (i.e., there are no transportation costs). Labor is homogeneous within a country but may have different productivities across countries. This implies that the production technology is assumed to differ across countries. Labor is costlessly mobile across industries within a country but is immobile across countries. Full employment of labor is also assumed. Consumers (the laborers) are assumed to maximize utility subject to an income constraint.

Below you will find a more complete description of each assumption along with a mathematical formulation of the model.

Perfect Competition

Perfect competition in all markets means that the following conditions are assumed to hold.

1. Many firms produce output in each industry such that each firm is too small for its output decisions to affect the market price. This implies that when choosing output to maximize profit, each firm takes the price as given or exogenous.
2. Firms choose output to maximize profit. The rule used by perfectly competitive firms is to choose the output level that equalizes the price (P) with the marginal cost (MC). That is, set $P = MC$.

3. Output is homogeneous across all firms. This means that goods are identical in all their characteristics such that a consumer would find products from different firms indistinguishable. We could also say that goods from different firms are perfect substitutes for all consumers.
4. There is free entry and exit of firms in response to profits. Positive profit sends a signal to the rest of the economy and new firms enter the industry. Negative profit (losses) leads existing firms to exit, one by one, out of the industry. As a result, in the long run economic profit is driven to zero in the industry.
5. Information is perfect. For example, all firms have the necessary information to maximize profit and to identify the positive profit and negative profit industries.

Two Countries

The case of two countries is used to simplify the model analysis. Let one country be the United States and the other France. Note that anything related exclusively to France in the model will be marked with an asterisk. The two countries are assumed to differ only with respect to the production technology.

Two Goods

Two goods are produced by both countries. We assume a barter economy. This means that no money is used to make transactions. Instead, for trade to occur, goods must be traded for other goods. Thus we need at least two goods in the model. Let the two produced goods be wine and cheese.

One Factor of Production

Labor is the one factor of production used to produce each of the goods. The factor is homogeneous and can freely move between industries.

Utility Maximization and Demand

In David Ricardo's original presentation of the model, he focused exclusively on the supply side. Only later did John Stuart Mill introduce demand into the model. Since much can be learned with Ricardo's incomplete model, we proceed initially without formally specifying demand or utility functions. Later in the chapter we will use the aggregate utility specification to depict an equilibrium in the model.

When needed, we will assume that aggregate utility can be represented by a function of the form $U = C_C C_W$, where C_C and C_W are the aggregate quantities of cheese and wine consumed in the country, respectively. This function is chosen because it has properties that make it easy to depict an equilibrium. The most important feature is that the function is homothetic, which implies that the country consumes wine and cheese in the same fixed proportion at given prices regardless of income. If two countries share the same homothetic preferences, then when the countries share the same prices, as they will in free trade, they will also consume wine and cheese in the same proportion.

General Equilibrium

The Ricardian model is a general equilibrium model. This means that it describes a complete circular flow of money in exchange for goods and services. Thus the sale of goods and services generates revenue to the firms that in turn is used to pay for the factor services (wages to workers in this case) used in production. The factor income (wages) is used, in turn, to buy the goods and services produced by the firms. This generates revenue to the firms and the cycle repeats again. A “general equilibrium” arises when prices of goods, services, and factors are such as to equalize supply and demand in all markets simultaneously.

Production

The production functions in [Table 2.4 "Production of Cheese"](#) and [Table 2.5 "Production of Wine"](#) represent industry production, not firm production. The industry consists of many small firms in light of the assumption of perfect competition.

Table 2.4 Production of Cheese

United States	France
$Q_C = \frac{L_C}{a_{LC}} \left[\frac{\text{hrs}}{\text{lb}} \right]$	$Q_C^* = \frac{L_C^*}{a_{LC}^*}$
where	
Q_C = quantity of cheese produced in the United States	

Chapter 2 The Ricardian Theory of Comparative Advantage

United States	France
L_C = amount of labor applied to cheese production in the United States	
a_{LC} = unit labor requirement in cheese production in the United States (hours of labor necessary to produce one unit of cheese)	
*All starred variables are defined in the same way but refer to the process in France.	

Table 2.5 Production of Wine

United States	France
$Q_W = \frac{L_W}{a_{LW}} \left[\frac{\text{hrs}}{\text{gal}} \right]$	$Q_W^* = \frac{L_W^*}{a_{LW}^*}$
where	
Q_W = quantity of wine produced in the United States	
L_W = amount of labor applied to wine production in the United States	
a_{LW} = unit labor requirement in wine production in the United States (hours of labor necessary to produce one unit of wine)	
*All starred variables are defined in the same way but refer to the process in France.	

The **unit labor requirements**⁵ define the technology of production in two countries. Differences in these labor costs across countries represent differences in technology.

5. The quantity of labor needed to produce one unit of a good.

Resource Constraint

The resource constraint in this model is also a labor constraint since labor is the only factor of production (see [Table 2.6 "Labor Constraints"](#)).

Table 2.6 Labor Constraints

United States	France
$L_C + L_W = L$	$L_{C^*} + L_{W^*} = L^*$

where

L = the labor endowment in the United States (the total number of hours the workforce is willing to provide)

When the resource constraint holds with equality, it implies that the resource is fully employed. A more general specification of the model would require only that the sum of labor applied in both industries be less than or equal to the labor endowment. However, the assumptions of the model will guarantee that production uses all available resources, and so we can use the less general specification with the equal sign.

Factor Mobility

The one factor of production, labor, is assumed to be immobile across countries. Thus labor cannot move from one country to another in search of higher wages. However, labor is assumed to be freely and costlessly mobile between industries within a country. This means that workers working in the one industry can be moved to the other industry without any cost incurred by the firms or the workers. The significance of this assumption is demonstrated in the immobile factor model in [Chapter 4 "Factor Mobility and Income Redistribution"](#).

Transportation Costs

The model assumes that goods can be transported between countries at no cost. This assumption simplifies the exposition of the model. If transport costs are included, it can be shown that the key results of the model may still be obtained.

Exogenous and Endogenous Variables

In describing any model, it is always useful to keep track of which variables are exogenous and which are endogenous. **Exogenous variables**⁶ are those variables in a model that are determined by processes that are not described within the model itself. When describing and solving a model, exogenous variables are taken as fixed parameters whose values are known. They are variables over which the agents within the model have no control. In the Ricardian model, the parameters (L , a_{LC} , a_{LW}) are exogenous. The corresponding starred variables are exogenous in the other country.

Endogenous variables⁷ are those variables determined when the model is solved. Thus finding the solution to a model means solving for the values of the endogenous variables. Agents in the model can control or influence the endogenous variables through their actions. In the Ricardian model, the variables (L_C , L_W , Q_C , Q_W) are endogenous. Likewise, the corresponding starred variables are endogenous in the other country.

KEY TAKEAWAYS

- The Ricardian model incorporates the standard assumptions of perfect competition.
- The simple Ricardian model assumes two countries producing two goods and using one factor of production.
- The goods are assumed to be identical, or homogeneous, within and across countries.
- The workers are assumed to be identical in the productive capacities within, but not across, countries.
- Workers can move freely and costlessly between industries but cannot move to another country.

6. A variable whose value is determined external to the model and whose value is known to the agents in the model. In the Ricardian model, the unit labor requirements and the labor endowment are exogenous.

7. A variable whose value is determined as an outcome of, or solution to, the model. In the Ricardian model, the allocation of workers to production, the quantities of the goods produced, and the terms of trade are endogenous.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The type of variable whose value is determined as a part of the solution to the model.
 - b. The type of variable whose value is determined outside the model and is presumed to be known by the model participants.
 - c. The rule used by perfectly competitive firms to determine the profit-maximizing level of output.
 - d. What a perfectly competitive firm may do if it experiences substantially negative profit.
 - e. The kind of equilibrium in a model in which multiple markets satisfy the equality of supply and demand simultaneously.
2. Suppose that the unit labor requirements for wine and cheese are $a_{LC} = 6$ hrs./lb. and $a_{LW} = 4$ hrs./gal., respectively, and that labor hours applied to cheese and wine production are 60 and 80, respectively. What is total output of cheese and wine?
3. Suppose that the unit labor requirements for wine and cheese are $a_{LC} = 3$ hrs./lb. and $a_{LW} = 2$ hrs./gal., respectively, and that labor hours applied to cheese and wine production are 60 and 80, respectively. What would the total output of wine be if all the labor hours were shifted to produce wine?

2.4 The Ricardian Model Production Possibility Frontier

LEARNING OBJECTIVE

1. Learn how the plot of the labor constraint yields the production possibility frontier.

Using the two production functions and the labor constraint, we can describe the **production possibility frontier (PPF)**⁸. First, note that the production functions can be rewritten as $L_C = a_{LC} Q_C$ and $L_W = a_{LW} Q_W$. Plugging these values for L_C and L_W into the labor constraint yields the equation for the PPF:

$$a_{LC} Q_C + a_{LW} Q_W = L.$$

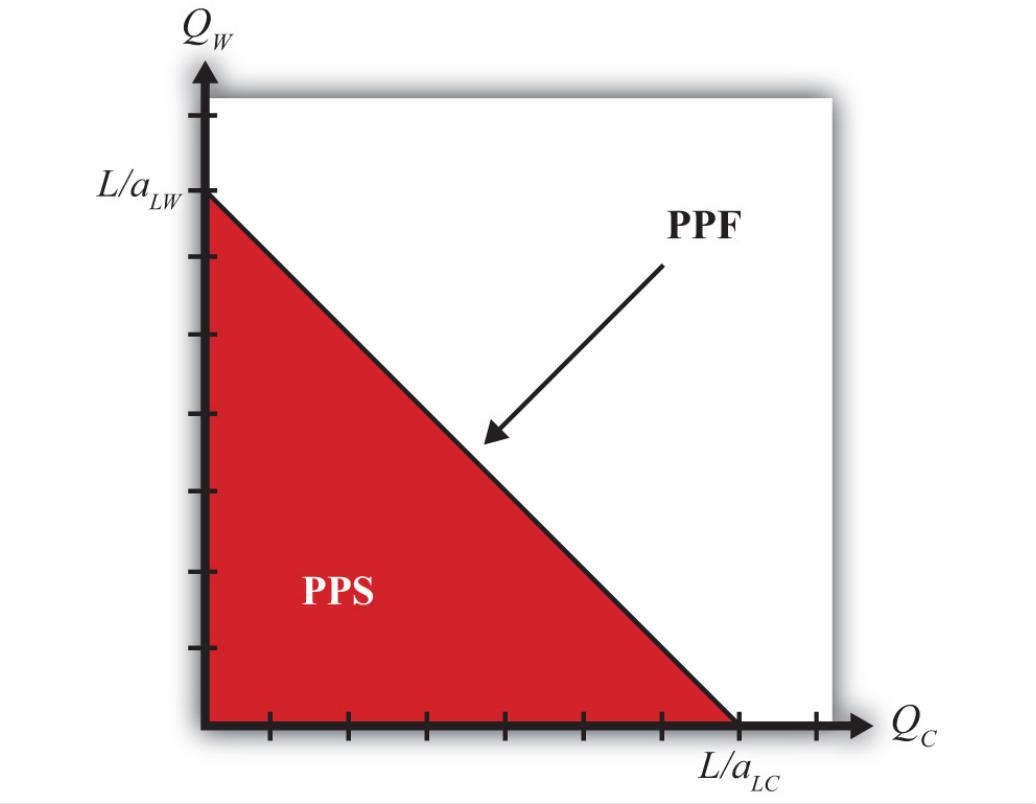
This equation has three exogenous variables (a_{LC} , a_{LW} , and L) that we assume have known values and two endogenous variables (Q_C and Q_W) whose values must be solved for. The PPF equation is a linear equation—that is, it describes a line. With some algebraic manipulation, we can rewrite the PPF equation into the standard form for an equation of a line, generally written as $y = mx + b$, where y is the variable on the vertical axis, x is the variable on the horizontal axis, m is the slope of the line, and b is the y -intercept. The PPF equation can be rewritten as

$$Q_W = \frac{L}{a_{LW}} - \left(\frac{a_{LC}}{a_{LW}} \right) Q_C.$$

We plot the PPF on the diagram in [Figure 2.1 "Production Possibilities"](#) with Q_C on the horizontal axis and Q_W on the vertical axis. The equation is easily plotted by following three steps.

8. The set of all output combinations that could be produced in a country when all the labor inputs are fully employed. In the Ricardian model, the PPF is linear.

Figure 2.1 Production Possibilities



1. Set $Q_C = 0$ and solve for Q_W . In this case, the solution is $Q_W = \frac{L}{a_{LW}}$. This corresponds to the Q_W -intercept. It tells us the quantity of wine that the United States could produce if it devoted all of its labor force (L) to the production of wine.
2. Set $Q_W = 0$ and solve for Q_C . In this case, the solution is $Q_C = \frac{L}{a_{LC}}$. This corresponds to the Q_C -intercept. It tells us the quantity of cheese that the United States could produce if it devoted all of its labor force (L) to the production of cheese.
3. Connect the two points with a straight line.

The straight downward-sloping line is the production possibility frontier. It describes all possible quantity combinations of wine and cheese that can be achieved by the U.S. economy. A movement along the curve represents a transfer of labor resources out of one industry and into another such that all labor remains employed.

Points inside the PPF are production possibilities but correspond to underemployment of labor resources. In fact, all production possibilities regardless

of whether full employment is fulfilled are referred to as the production possibility set (PPS). The PPS is represented by all the points within and on the border of the red triangle in [Figure 2.1 "Production Possibilities"](#).

KEY TAKEAWAYS

- The equation $a_{LC} Q_C + a_{LW} Q_W = L$ is an equation of a line whose plot represents the country's production possibility frontier (PPF).
- A PPF is the combination of outputs of cheese and wine that the country can produce given a production technology (i.e., given that unit labor requirements are exogenous) and assuming all of its labor hours are employed.
- A production possibility set (PPS) is the combination of outputs that a country can produce even if some of the labor is unemployed.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term describing the set of all output combinations that can be produced within an economy.
 - b. The term describing the set of all output combinations that can be produced within an economy with full employment of all available resources.
2. Suppose that the unit labor requirements for wine and cheese are $a_{LC} = 6$ hrs./lb., $a_{LW} = 4$ hrs./gal., respectively, and that total labor hours available for production are 60. What is the maximum output of cheese? What is the maximum output of wine?
3. Suppose that the unit labor requirements for wine and cheese are $a_{LC} = 6$ hrs./lb. and $a_{LW} = 4$ hrs./gal., respectively, and that total labor hours available for production are 60. Plot the production possibility frontier.

2.5 Definitions: Absolute and Comparative Advantage

LEARNING OBJECTIVES

1. Learn how to define labor productivity and opportunity cost within the context of the Ricardian model.
2. Learn to identify and distinguish absolute advantage and comparative advantage.
3. Learn to identify comparative advantage via two methods: (1) by comparing opportunity costs and (2) by comparing relative productivities.

The basis for trade in the Ricardian model is differences in technology between countries. Below we define two different ways to describe technology differences. The first method, called absolute advantage, is the way most people understand technology differences. The second method, called comparative advantage, is a much more difficult concept. As a result, even those who learn about comparative advantage often will confuse it with absolute advantage. It is quite common to see misapplications of the principle of comparative advantage in newspaper and journal stories about trade. Many times authors write “comparative advantage” when in actuality they are describing absolute advantage. This misconception often leads to erroneous implications, such as a fear that technology advances in other countries will cause our country to lose its comparative advantage in everything. As will be shown, this is essentially impossible.

To define absolute advantage, it is useful to define labor productivity first. To define comparative advantage, it is useful to first define opportunity cost. Next, each of these is defined formally using the notation of the Ricardian model.

Labor Productivity

Labor productivity⁹ is defined as the quantity of output that can be produced with a unit of labor. Since a_{LC} represents hours of labor needed to produce one pound of cheese, its reciprocal, $1/a_{LC}$, represents the labor productivity of cheese production in the United States. Similarly, $1/a_{LW}$ represents the labor productivity of wine production in the United States.

9. The quantity of a good that can be produced per unit of labor input. It is the reciprocal of the unit labor requirement.

Absolute Advantage

A country has an **absolute advantage**¹⁰ in the production of a good relative to another country if it can produce the good at lower cost or with higher productivity. Absolute advantage compares industry productivities across countries. In this model, we would say the United States has an absolute advantage in cheese production relative to France if

$$a_{LC} < a_{LC}^*$$

or if

$$\frac{1}{a_{LC}} > \frac{1}{a_{LC}^*}.$$

The first expression means that the United States uses fewer labor resources (hours of work) to produce a pound of cheese than does France. In other words, the resource cost of production is lower in the United States. The second expression means that labor productivity in cheese in the United States is greater than in France. Thus the United States generates more pounds of cheese per hour of work.

Obviously, if $a_{LC}^* < a_{LC}$, then France has the absolute advantage in cheese. Also, if $a_{LW} < a_{LW}^*$, then the United States has the absolute advantage in wine production relative to France.

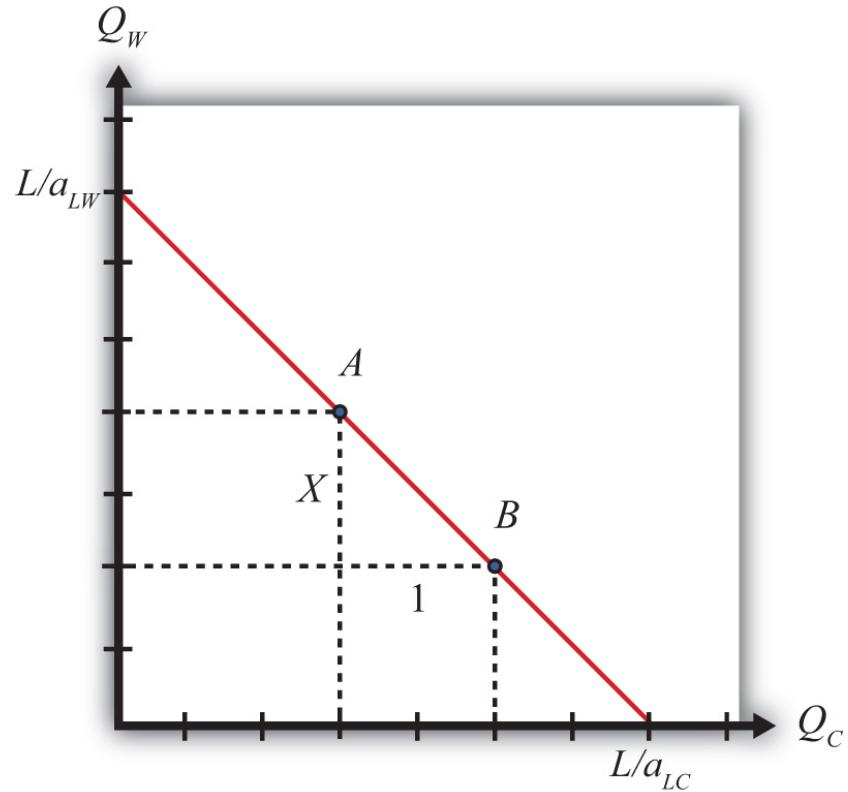
Opportunity Cost

Opportunity cost¹¹ is defined generally as the value of the next best opportunity. In the context of national production, the nation has opportunities to produce wine and cheese. If the nation wishes to produce more cheese, then because labor resources are scarce and fully employed, it is necessary to move labor out of wine production in order to increase cheese production. The loss in wine production necessary to produce more cheese represents the opportunity cost to the economy. The slope of the PPF, $-(a_{LC}/a_{LW})$, corresponds to the opportunity cost of production in the economy.

10. A country has an absolute advantage in the production of a good if it can produce the good at a lower labor cost and if labor productivity in the good is higher than in another country.

11. The value or quantity of something that must be given up to obtain something else. In the Ricardian model, opportunity cost is the amount of a good that must be given up to produce one more unit of another good.

Figure 2.2 Defining Opportunity Cost



To see this more clearly, consider points A and B in [Figure 2.2 "Defining Opportunity Cost"](#). Let the horizontal distance between A and B be one pound of cheese. Label the vertical distance X. The distance X then represents the quantity of wine that must be given up to produce one additional pound of cheese when moving from point A to B. In other words, X is the opportunity cost of producing cheese.

Note also that the slope of the line between A and B is given by the formula

$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{-X}{1}.$$

Thus the slope of the line between A and B is the opportunity cost, which from above is given by $-(a_{LC}/a_{LW})$. We can more clearly see why the slope of the PPF represents the opportunity cost by noting the units of this expression:

$$\frac{-a_{LC}}{a_{LW}} \left[\frac{\frac{\text{hrs}}{\text{lb}}}{\frac{\text{hrs}}{\text{gal}}} = \frac{\text{gal}}{\text{lb}} \right].$$

Thus the slope of the PPF expresses the number of gallons of wine that must be given up (hence the minus sign) to produce another pound of cheese. Hence it is the opportunity cost of cheese production (in terms of wine). The reciprocal of the slope, $-(a_{LW}/a_{LC})$, in turn represents the opportunity cost of wine production (in terms of cheese).

Since in the Ricardian model the PPF is linear, the opportunity cost is the same at all possible production points along the PPF. For this reason, the Ricardian model is sometimes referred to as a *constant (opportunity) cost model*.

Comparative Advantage Using Opportunity Costs

A country has a comparative advantage in the production of a good if it can produce that good at a lower opportunity cost relative to another country. Thus the United States has a comparative advantage in cheese production relative to France if

$$\frac{a_{LC}}{a_{LW}} < \frac{a_{LC}^*}{a_{LW}^*}.$$

This means that the United States must give up less wine to produce another pound of cheese than France must give up to produce another pound. It also means that the slope of the U.S. PPF is flatter than the slope of France's PPF.

Starting with the inequality above, cross multiplication implies the following:

$$\frac{a_{LC}}{a_{LW}} < \frac{a_{LC}^*}{a_{LW}^*} \Rightarrow \frac{a_{LW}^*}{a_{LC}^*} < \frac{a_{LW}}{a_{LC}}.$$

This means that France can produce wine at a lower opportunity cost than the United States. In other words, France has a comparative advantage in wine production. This also means that if the United States has a comparative advantage in one of the two goods, France must have the comparative advantage in the other good. It is not possible for one country to have the comparative advantage in both of the goods produced.

Suppose one country has an absolute advantage in the production of both goods. Even in this case, each country will have a comparative advantage in the production of one of the goods. For example, suppose $a_{LC} = 10$, $a_{LW} = 2$, $a_{LC}^* = 20$, and $a_{LW}^* = 5$. In this case, $a_{LC} (10) < a_{LC}^* (20)$ and $a_{LW} (2) < a_{LW}^* (5)$, so the United States has the absolute advantage in the production of both wine and cheese. However, it is also true that

$$\frac{a_{LC}^*}{a_{LW}^*} \left(\frac{20}{5} \right) < \frac{a_{LC}}{a_{LW}} \left(\frac{10}{2} \right)$$

so that France has the comparative advantage in cheese production relative to the United States.

Using Relative Productivities

Another way to describe comparative advantage is to look at the relative productivity advantages of a country. In the United States, the labor productivity in cheese is $1/10$, while in France it is $1/20$. This means that the U.S. productivity advantage in cheese is $(1/10)/(1/20) = 2/1$. Thus the United States is twice as productive as France in cheese production. In wine production, the U.S. advantage is $(1/2)/(1/5) = (2.5)/1$. This means the United States is two and one-half times as productive as France in wine production.

The comparative advantage good in the United States, then, is that good in which the United States enjoys the greatest productivity advantage: wine.

Also consider France's perspective. Since the United States is two times as productive as France in cheese production, then France must be $1/2$ times as productive as the United States in cheese. Similarly, France is $2/5$ times as productive in wine as the United States. Since $1/2 > 2/5$, France has a disadvantage in production of both goods. However, France's disadvantage is smallest in cheese; therefore, France has a comparative advantage in cheese.

No Comparative Advantage

The only case in which neither country has a comparative advantage is when the opportunity costs are equal in both countries. In other words, when

$$\frac{a_{LC}}{a_{LW}} = \frac{a_{LC}^*}{a_{LW}^*},$$

then neither country has a comparative advantage. It would seem, however, that this is an unlikely occurrence.

KEY TAKEAWAYS

- Labor productivity is defined as the quantity of output produced with one unit of labor; in the model, it is derived as the reciprocal of the unit labor requirement.
- Opportunity cost is defined as the quantity of a good that must be given up in order to produce one unit of another good; in the model, it is defined as the ratio of unit labor requirements between the first and the second good.
- The opportunity cost corresponds to the slope of the country's production possibility frontier (PPF).
- An absolute advantage arises when a country has a good with a lower unit labor requirement and a higher labor productivity than another country.
- A comparative advantage arises when a country can produce a good at a lower opportunity cost than another country.
- A comparative advantage is also defined as the good in which a country's relative productivity advantage (disadvantage) is greatest (smallest).
- It is not possible that a country does not have a comparative advantage in producing something unless the opportunity costs (relative productivities) are equal. In this case, neither country has a comparative advantage in anything.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The labor productivity in cheese if four hours of labor are needed to produce one pound.
 - b. The labor productivity in wine if three kilograms of cheese can be produced in one hour and ten liters of wine can be produced in one hour.
 - c. The term used to describe the amount of labor needed to produce a ton of steel.
 - d. The term used to describe the quantity of steel that can be produced with an hour of labor.
 - e. The term used to describe the amount of peaches that must be given up to produce one more bushel of tomatoes.
 - f. The term used to describe the slope of the PPF when the quantity of tomatoes is plotted on the horizontal axis and the quantity of peaches is on the vertical axis.
2. Consider a Ricardian model with two countries, the United States and Ecuador, producing two goods, bananas and machines. Suppose the unit labor requirements are $a_{LB}^{US} = 8$, $a_{LB}^E = 4$, $a_{LM}^{US} = 2$, and $a_{LM}^E = 4$. Assume the United States has 3,200 workers and Ecuador has 400 workers.
 - a. Which country has the absolute advantage in bananas? Why?
 - b. Which country has the comparative advantage in bananas? Why?
 - c. How many bananas and machines would the United States produce if it applied half of its workforce to each good?
3. Consider a Ricardian model with two countries, England and Portugal, producing two goods, wine and corn. Suppose the unit labor requirements in wine production are $a_{LW}^{Eng} = 1/3$ hour per liter and $a_{LW}^{Port} = 1/2$ hour per liter, while the unit labor requirements in corn are $a_{LC}^{Eng} = 1/4$ hour per kilogram and $a_{LC}^{Port} = 1/2$ hour per kilogram.

- a. What is labor productivity in the wine industry in England and in Portugal?
- b. What is the opportunity cost of corn production in England and in Portugal?
- c. Which country has the absolute advantage in wine? In corn?
- d. Which country has the comparative advantage in wine? In corn?

2.6 A Ricardian Numerical Example

LEARNING OBJECTIVES

1. Using a numerical example similar to one used by David Ricardo, learn how specialization in one's comparative advantage good can raise world productive efficiency.
2. Learn how both countries can consume more of both goods after trade.

The simplest way to demonstrate that countries can gain from trade in the Ricardian model is by use of a numerical example. This is how Ricardo presented his argument originally. The example demonstrates that both countries will gain from trade if they specialize in their comparative advantage good and trade some of it for the other good. We set up the example so that one country (the United States) has an absolute advantage in the production of both goods. Ricardo's surprising result was that a country can gain from trade even if it is technologically inferior in producing every good. Adam Smith explained in *The Wealth of Nations* that trade is advantageous to both countries, but in his example each country had an absolute advantage in one of the goods. That trade could be advantageous if each country specializes in the good in which it has the technological edge is not surprising at all.

Suppose the exogenous variables in the two countries take the values in [Table 2.7 "Exogenous Variable Values"](#).

Table 2.7 Exogenous Variable Values

United States	$a_{LC} = 1$	$a_{LW} = 2$	$L = 24$
France	$a_{LC*} = 6$	$a_{LW*} = 3$	$L* = 24$
where			
L = the labor endowment in the United States (the total number of hours the workforce is willing to provide)			

a_{LC} = unit labor requirement in cheese production in the United States (hours of labor necessary to produce one unit of cheese)

a_{LW} = unit labor requirement in wine production in the United States (hours of labor necessary to produce one unit of wine)

* All starred variables are defined in the same way but refer to the process in France.

By assumption, the United States has the absolute advantage in cheese production and wine production because $a_{LC}(1) < a_{LC}^*(6)$ and $a_{LW}(2) < a_{LW}^*(3)$.

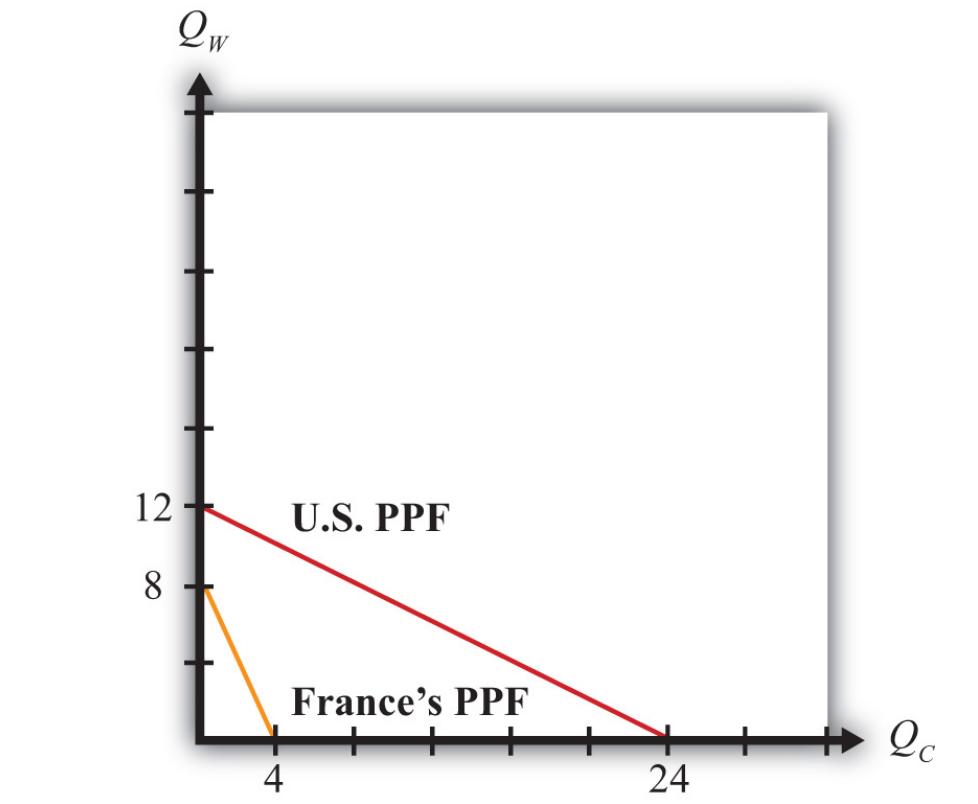
The United States also has the comparative advantage in cheese production because $\frac{a_{LC}}{a_{LW}} \left(\frac{1}{2}\right) < \frac{a_{LC}^*}{a_{LW}^*} \left(\frac{6}{3}\right)$. The cost of producing cheese in the United States is one half gallon of wine per pound of cheese. In France, it is two gallons per pound.

France, however, has the comparative advantage in wine production because $\frac{a_{LW}^*}{a_{LC}^*} \left(\frac{3}{6}\right) < \frac{a_{LW}}{a_{LC}} \left(\frac{2}{1}\right)$. The cost of producing wine in France is one half pound of cheese per gallon of wine, while in the United States, it is two pounds per gallon.

The production possibility frontiers for both countries are plotted on [Figure 2.3 "Production Possibility Frontiers"](#). Notice that the U.S. PPF lies outside France's PPF. Since both countries are assumed to be the same size in the example, this indicates the U.S. absolute advantage in the production of both goods.

The absolute value of the slope of each PPF represents the opportunity cost of cheese production. Since the U.S. PPF is flatter than France's, this means that the opportunity cost of cheese production is lower in the United States and thus indicates that the United States has the comparative advantage in cheese production.

Figure 2.3 *Production Possibility Frontiers*



With full employment of labor, production will occur at some point along the PPF.

To see the effects of specialization and free trade, we must compare it to a situation of no trade, or autarky. Thus we must construct an autarky equilibrium first. To determine the autarky production point requires some information about the consumer demand for the goods. Producers will produce whatever consumers demand at the prevailing prices such that supply of each good equals demand. In autarky, this means that the production and consumption point for a country are the same.

For the purpose of this example, we will simply make up a plausible production and consumption point under autarky. Essentially, we assume that consumer demands are such as to generate the chosen production point. [Table 2.8 "Autarky Production and Consumption"](#) shows the autarky production and consumption levels for the two countries. It also shows total world production for each of the goods.

Table 2.8 Autarky Production and Consumption

	Cheese (lbs.)	Wine (gals.)
United States	16	4
France	3	2
World Total	19	6

Autarky Production and Consumption Points

In [Figure 2.4 "Autarky Equilibriums"](#) we depict the autarky production and consumption points for the United States and France. Each point lies on the interior section of the country's production possibility frontier.

Question: How do you know that the chosen production points are on the country's PPF?

Answer: To verify that a point is on the PPF, we can simply plug the quantities into the PPF equation to see if it is satisfied. The PPF formula is $a_{LC}Q_C + a_{LW}Q_W = L$. If we plug the exogenous variables for the United States into the formula, we get $Q_C + 2Q_W = 24$. Plugging in the production point from [Table 2.8 "Autarky Production and Consumption"](#) yields $16 + 2(4) = 24$, and since $16 + 8 = 24$, the production point must lie on the PPF.

Ricardo argued that trade gains could arise if countries first specialized in their comparative advantage good and then traded with the other country. Specialization in the example means that the United States produces only cheese and no wine, while France produces only wine and no cheese. These quantities are shown in [Table 2.9 "Production with Specialization in the Comparative Advantage Good"](#). Also shown are the world totals for each of the goods.

Figure 2.4 Autarky Equilibria

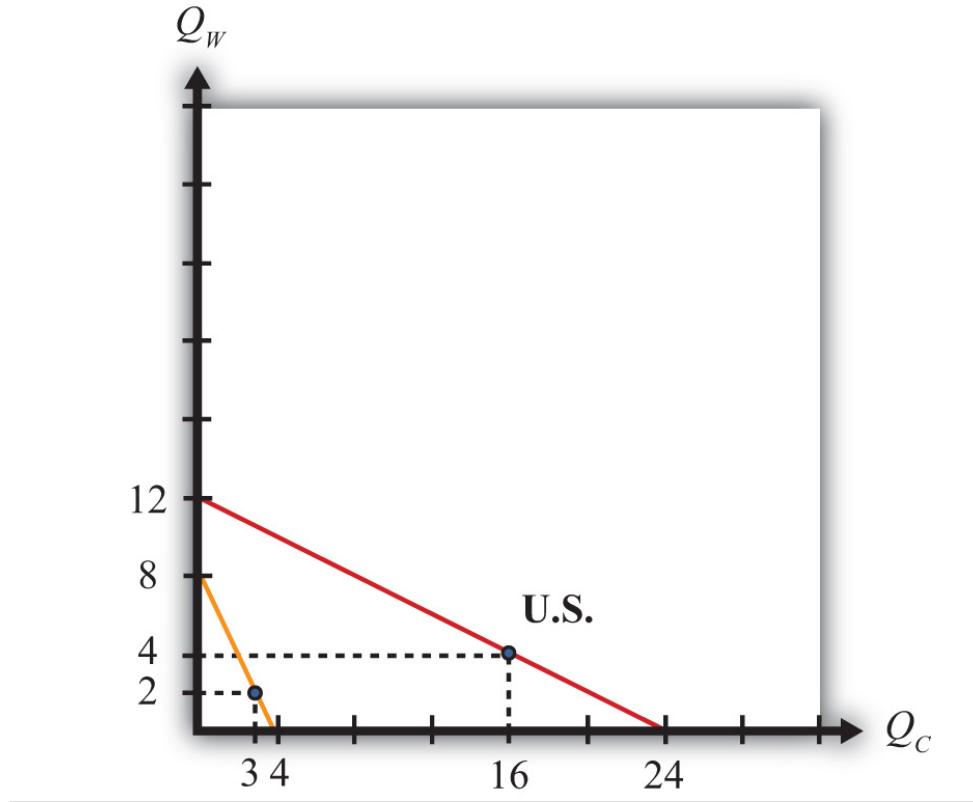


Table 2.9 Production with Specialization in the Comparative Advantage Good

	Cheese (lbs.)	Wine (gals.)
United States	24	0
France	0	8
World Total	24	8

At this point, we can already see a remarkable result. When countries specialize in their comparative advantage good, world output of both wine and cheese rises. Cheese output rises from nineteen to twenty-four pounds. Wine output rises from six to eight gallons. What's more, the output increases occur without an increase in the quantity of labor used to produce them. In autarky, it took forty-eight worker hours to produce nineteen pounds of cheese and six gallons of wine. With specialization, the same forty-eight worker hours produce twenty-four pounds of cheese and eight gallons of wine. This means that there is an increase in world productivity—more output per unit of labor. Often this productivity improvement is referred to as an increase or improvement in world production efficiency.

The increase in world production efficiency does not benefit the countries unless they can trade with each other after specialization. Both production points were feasible under autarky, but the countries demanded some of each good. Thus the countries will want some of each good after specialization, and the only way to accomplish this is through trade. Now if the world can produce more of both goods through specialization, clearly there must be a way to divide the surplus between the two countries so that each country ends up with more of both goods after trade than it had in autarky.

The surplus in world production amounts to five extra pounds of cheese and two extra gallons of wine. To assure that trade is advantageous for the two countries, each must have at least as much to consume of one good and more to consume of the other. Suppose we split the wine surplus equally and give three extra pounds of cheese to France and two extra pounds to the United States. Since the United States consumed sixteen pounds of cheese and four gallons of wine in autarky, it would now have eighteen pounds of cheese and five gallons of wine after specialization and trade. France, which began with three pounds of cheese and two gallons of wine in autarky, would now have six pounds of cheese and three gallons of wine. Consumption and production after trade for the two countries is shown in [Table 2.10 "Consumption and Production after Trade"](#).

Table 2.10 Consumption and Production after Trade

Country	Cheese (lbs.)		Wine (gals.)	
	Consumption	Production	Consumption	Production
United States	18	24	5	0
France	6	0	3	8
World Total	24	24	8	8

In order for consumption of both goods to be higher in both countries, trade must occur. In the example, the United States is consuming five gallons of wine and producing none, so it must import the five gallons from France. France is consuming six pounds of cheese with no cheese production, so it must import the six pounds from the United States. The terms of trade is $TOT = 5 \text{ gals.}/6 \text{ lbs.}$, or $5/6 \text{ gals./lb.}$

Exercise Conclusion

The Ricardian model numerical example assumes that countries differ in their production technologies such that one of the countries is absolutely more productive than the other in the production of each of the two goods. If these two countries specialize in their comparative advantage good, then world production rises for both goods. Increased output occurs even though there is no increase in the amount of labor input in the world; thus the example demonstrates that specialization can raise world production efficiency. Because of the increase in output, it is possible to construct a terms of trade between the countries such that each country consumes more of each good with specialization and trade than was possible under autarky. Thus both countries can gain from trade. The surprising result of this example is that a country that is technologically inferior to another in the production of all goods can nevertheless benefit from trade with that country.

Limitations of the Numerical Example

A numerical example can display only one possible outcome for the model. As such, all conclusions should be viewed as possibilities rather than general results of the model. With further thought, there are some problems with the example. First, it is conceivable that with a different choice for the country's autarky production and consumption points, world output might not rise for both goods upon specialization. In this case, we could not be sure that both countries would gain from trade. Second, since we merely made up a terms of trade that generated the interesting conclusion, we could ask whether a favorable terms of trade is likely to arise. Is it possible to make up a different terms of trade such that one country enjoys all the benefits of increased production while the other is made worse off? How can we be sure that this outcome would not arise? Finally, even if the country has more of both goods after trade, can we be sure that all consumers would have more of both goods? Perhaps some consumers would have more and others less.

The answer to some of these questions can be found by describing more carefully some of the features of the model. In particular, we must describe the relationship between prices and wages. Using these relationships, we can explain the impact of free trade on the price ratio and the effect of trade on the distribution of income.

KEY TAKEAWAYS

- In a two-country, two-good, one-factor Ricardian model, specialization in each country's comparative advantage good can raise world output of both goods.
- An increase in world output given the same level of inputs is called an increase in world productive efficiency.
- By choosing an appropriate terms of trade, both countries can consume more of both goods relative to autarky.

EXERCISE

1. Consider a Ricardian model with two countries, the United States and the EU, producing two goods, soap bars and toothbrushes. Suppose the productivities are $a_{LS}^{US} = 2$ soap bars per worker, $a_{LS}^E = 4$ soap bars per worker, $a_{LT}^{US} = 8$ toothbrushes per worker, and $a_{LT}^E = 4$ toothbrushes per worker. Assume the United States has 3,200 workers and the EU has 4,000 workers.
 - a. Plot the PPFs for both countries.
 - b. Determine how much each country would produce if it specialized in its comparative advantage good.
 - c. Now choose a plausible autarky production point on each country's PPF such that the world output of each good is exceeded by the outputs determined in part b.
 - d. Determine a terms of trade between the two countries that will assure that both countries can consume more of both goods after trade.

2.7 Relationship between Prices and Wages

LEARNING OBJECTIVE

1. Learn how worker wages and the prices of the goods are related to each other in the Ricardian model.

The Ricardian model assumes that the wine and cheese industries are both perfectly competitive. Among the assumptions of perfect competition is free entry and exit of firms in response to economic profit. If positive profits are being made in one industry, then because of perfect information, profit-seeking entrepreneurs will begin to open more firms in that industry. The entry of firms, however, raises industry supply, which forces down the product price and reduces profit for every other firm in the industry. Entry continues until economic profit is driven to zero. The same process occurs in reverse when profit is negative for firms in an industry. In this case, firms will close down one by one as they seek more profitable opportunities elsewhere. The reduction in the number of firms reduces industry supply, which raises the product's market price and raises profit for all remaining firms in the industry. Exit continues until economic profit is raised to zero. This implies that if production occurs in an industry, be it in autarky or free trade, then economic profit must be zero.

Profit is defined as total revenue minus total cost. Let Π_C represent profit in the cheese industry. We can write this as

$$\Pi_C = P_C Q_C - w_C L_C = 0,$$

where P_C is the price of cheese in dollars per pound, w_C is the wage paid to workers in dollars per hour, $P_C Q_C$ is total industry revenue, and $w_C L_C$ is total industry cost. By rearranging the zero-profit condition, we can write the wage as a function of everything else to get

$$w_C = \frac{P_C Q_C}{L_C}.$$

Recall that the production function for cheese is $Q_C = \frac{L_C}{a_{LC}}$. Plugging this in for Q_C above yields

$$w_C = \frac{P_C \left(\frac{L_C}{a_{LC}} \right)}{L_C} = \frac{P_C}{a_{LC}}$$

or just

$$w_C = \frac{P_C}{a_{LC}}.$$

If production occurs in the wine industry, then profit will be zero as well. By the same algebra we can get

$$w_W = \frac{P_W}{a_{LW}}.$$

KEY TAKEAWAYS

- The assumption of free entry and exit in perfect competition implies that industry profit will be zero when the market is in equilibrium.
- Nominal wages (meaning wages measured in dollars) to workers in each industry will equal the output price divided by the unit labor requirement in that industry.

EXERCISE

1. Starting with the zero-profit condition in the wine industry, show why the winemaker's wage depends on the price of wine and wine productivity.

2.8 Deriving the Autarky Terms of Trade

LEARNING OBJECTIVES

1. Learn how the autarky terms of trade is determined in a Ricardian model.
2. Learn why free and costless labor mobility and homogeneous labor force wages to be equal in both industries.

The Ricardian model assumes that all workers are identical, or homogeneous, in their productive capacities and that labor is freely mobile across industries. In autarky, assuming at least one consumer demands some of each good, the country will produce on the interior of its PPF. That is, it will produce some wine and some cheese.

Question: Profit-maximizing firms would never set a wage rate above the level set in the other industry. Why?

Answer: Suppose the cheese industry set a higher wage such that $w_C > w_W$. In this case, all the wine workers would want to move to the cheese industry for any wage greater than w_W . Since their productivity in cheese is the same as the current cheese workers and since it does not cost anything for them to move to the other industry, the cheese industry could lower their costs and raise profit by paying a lower wage. To maximize profit, they must lower their wage. Thus only equal wage rates can be sustained between two perfectly competitive producing industries in the Ricardian model.

In autarky, then, $w_C = w_W$. Plugging in the relationships derived in the previous section yields

$$\frac{P_W}{a_{LW}} = \frac{P_C}{a_{LC}}$$

or

$$\left(\frac{P_C}{P_W} \right)^{Aut} = \frac{a_{LC}}{a_{LW}} .$$

This means that the autarky price ratio (cheese over wine) or terms of trade equals the opportunity cost of producing cheese. Another way to say the same thing is that the price of cheese (in terms of wine) in autarky equals the opportunity cost of producing cheese (in terms of wine).

Question: Why is there an autarky terms of trade when there is no trade in autarky?

Answer: The Ricardian model represents a barter economy. Even though we define prices and wages in monetary terms, all relevant solutions in the model are described in terms of ratios in which the money or dollars cancel out. Never will we solve explicitly for the dollar price of wine or cheese or the dollar wage rate.

Thus a good way to think about how the model works is to imagine that workers go to work in their respective industries and produce wine or cheese. At the end of the day, they are paid not in dollars but in goods. The cheese workers' wage is a quantity of cheese. The wine workers earn a quantity of wine. Since workers, as consumers, presumably will desire some wine and some cheese for their evening dinner, they must first go to a market to trade some of their wages (goods) for some of the other goods available at the market.

In autarky, cheese workers and wine workers come together on the domestic market to trade their goods. The autarky price ratio or terms of trade represents the amount of wine that exchanges per unit of cheese on the domestic barter market.

KEY TAKEAWAY

- The autarky terms of trade (cheese in terms of wine) equals the opportunity cost (of cheese in terms of wine).

EXERCISE

1. Use the information below to answer the following questions.

TABLE 2.11 LABOR PRODUCTIVITY IN ITALY
AND GERMANY

	Beer	Pizza
Italian Labor Productivity	6 bottles/hour	6 pizzas/hour
German Labor Productivity	5 bottles/hour	3 pizzas/hour

- a. Which country has the absolute advantage in beer? In pizza? Explain why.
- b. Explain why Italy's comparative advantage good is the one it can produce "most better," while Germany's comparative advantage good is the one it can produce "least worse."
- c. What autarky price ratios (P_B/P_P) would prevail in each country? Explain. Be sure to include units.

2.9 The Motivation for International Trade and Specialization

LEARNING OBJECTIVES

1. Learn that differences in autarky prices (terms of trade) coupled with the profit-seeking motive and the absence of transportation costs induce international trade.
2. Learn how the price changes that occur with trade induce specialization.

The Ricardian model can be used to explain Adam Smith's *invisible hand*. The invisible hand refers to the ability of the market, or the market mechanism, to allocate resources to their best possible uses. In the presentation of the Ricardian model it seems as if one must apply a mathematical formula (comparing opportunity costs) to identify which country has a comparative advantage and then instruct firms (perhaps by government decree) as to which goods they ought to produce.

Fortunately, none of this is necessary if the market, or the invisible hand, is allowed to operate. Instead, firms, or their owners, motivated entirely by profit, would automatically choose the appropriate good to produce and trade. In so doing, they would be led to maximize the output of goods and satisfy consumer demands to the extent possible given the limited resources in the economy. In *The Wealth of Nations*, Adam Smith said, “[An individual is] led by an *invisible hand* to promote an end which was no part of his intention.” See Book 4, Chapter 2 in Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, McMaster University Archive for the History of Economic Thought, <http://socserv2.socsci.mcmaster.ca/~econ/ugcm/3ll3.smith/wealth/wealbk04>. Emphasis mine. Maximizing society’s welfare is not the profit seeker’s intention; instead, he intends only to do what is best for himself. However, by virtue of the wonders of the market mechanism, everyone is made better off as well. Here’s how it works in this context.

The Market Motivation to Trade

Suppose two countries, the United States and France, are initially in autarky. Assume the United States has a comparative advantage in cheese production relative to France. This implies

$$\frac{a_{LC}}{a_{LW}} < \frac{a_{LC}^*}{a_{LW}^*} .$$

This, in turn, implies

$$\left(\frac{P_C}{P_W} \right)^{Aut} < \left(\frac{P_C^*}{P_W^*} \right)^{Aut}.$$

This means that the autarky price of cheese in France (in terms of wine) is greater than the autarky price of cheese in the United States. In other words, you can buy more wine with a pound of cheese in the French market than you can in the U.S. market.

Similarly, by rearranging the above inequality,

$$\left(\frac{P_W}{P_C} \right)^{Aut} > \left(\frac{P_W^*}{P_C^*} \right)^{Aut},$$

which means that the autarky price of wine is higher in the United States (in terms of cheese) than it is in France. In other words, a gallon of wine can be exchanged for more cheese in the United States than it will yield in the French market.

Next, suppose the barriers to trade that induced autarky are suddenly lifted and the United States and France are allowed to trade freely. For simplicity, we assume there are no transportation costs to move the products across borders.

Differences in price ratios between countries and the desire to make more profit are sufficient to generate international trade. To explain why, it is useful to incorporate some friction in the trading process and to tell a dynamic story about how a new free trade equilibrium is reached.

First, note that the higher price of cheese in France means that cheese workers in the United States could get more wine for their cheese in France than in the United States. Suppose one by one over time cheese workers begin to take advantage of the opportunity for trade and begin to sell their cheese in the French market. We assume that some workers are more internationally adroit and thus move first. The motivation here is profit. Workers want to get more for the goods they are selling. As the U.S. cheese workers appear in the French market, the supply of cheese increases. This also represents exports of cheese from the United States to France. The increased supply will reduce the price of cheese in the French market, meaning that over time, the quantity of wine obtained for a pound of cheese will fall. Thus P_C^*/P_W^* falls once trade is opened.

Next, consider French wine workers immediately after trade opens. Since the price of wine is higher in the United States, French wine workers will one by one over time begin to sell their wine in the U.S. market. This represents exports of wine from France to the United States. The increased supply of wine to the United States lowers its price on the U.S. market. Thus each gallon of wine will trade for less and less cheese. This means P_W/P_C falls, which also means that its reciprocal, P_C/P_W , rises.

These shifts in supply will continue as long as the prices for the goods continue to differ between the two markets. Once the prices are equalized, there will be no incentive to trade any additional amount. Equalized prices mean that a pound of cheese will trade for the same number of gallons of wine in both markets. The free trade prices will be those prices that equalize total supply of each good in the world with total demand for each good.

As a result of trade, the price ratio, or terms of trade, will lie in between the two countries' autarky price ratios. In other words, the following inequality will result:

$$\left(\frac{P_C}{P_W}\right)^{\text{Aut}} < \left(\frac{P_C}{P_W}\right)^{\text{FT}} < \left(\frac{P_C^*}{P_W^*}\right)^{\text{Aut}}.$$

Whether the free trade price ratio will be closer to the U.S. or France's autarky price ratio will depend on the relative demands of cheese to wine in the two countries. These demands in turn will depend on the size of the countries. If the United States is a much larger country, in that it has a larger workforce, it will have a larger demand for both wine and cheese. When trade opens, the addition of France's supply and demand will have a relatively small effect on the U.S. price. Thus the free trade price ratio will be closer to the U.S. autarky price ratio.

The Market Motivation for Specialization

Once the prices begin to change because of trade, they will also affect the profitability of producing the two goods. In the United States, the price of cheese, its export good, will rise in moving to trade, while the price of wine, its import good, will fall. As shown above, the final price ratio in the United States (cheese to wine) in free trade will be greater than the autarky price ratio, so that

$$\left(\frac{P_C}{P_W}\right)^{\text{FT}} > \left(\frac{P_C}{P_W}\right)^{\text{Aut}}.$$

Because the autarky price ratio equals the opportunity cost of cheese production, it follows that

$$\left(\frac{P_C}{P_W} \right)^{\text{FT}} > \frac{a_{LC}}{a_{LW}}.$$

Note that this inequality will be true as soon as the price deviates from the autarky price and long before the free trade prices are reached. This also means that shortly after trade begins, the price of cheese (measured in terms of wine) exceeds the cost of producing cheese (also measured in terms of wine). Normally, when we measure the price and cost in dollar terms, when the price per unit exceeds the cost per unit, then positive profit is realized. The same is true when we measure the price and cost in terms of wine. Thus as soon as trade begins to change prices, cheese production becomes more profitable in the United States. And because we assume people are profit seeking, they will therefore seek to expand cheese production. But where will they find the workers to do so? There is only one place: wine workers. To expand cheese production, the country will have to give up wine production. But why do that?

Well, when the price of cheese in terms of wine exceeds the opportunity cost of cheese, it is also true, via cross multiplication, that

$$\frac{a_{LW}}{a_{LC}} > \left(\frac{P_W}{P_C} \right)^{\text{FT}}.$$

This means that the cost of producing wine (in terms of cheese) exceeds the price of wine (also in terms of cheese). Because cost is greater than price, profit is negative in the wine industry in the United States. That means wine producers have an incentive to shut down. And when they do, those workers can be moved into the cheese industry, where profit seekers wish to expand.

Thus, as long as individuals are profit seeking, the price differences that arise in autarky will be sufficient to induce export and specialization in the comparative advantage good. There is no need to use the complicated opportunity cost formula to first identify the comparative advantage good and no need to tell anyone what to do. Instead, the free market mechanism—Adam Smith’s invisible hand—is all that it takes.

KEY TAKEAWAYS

- A country with the lower price for a good in terms of the other good and compared to the other country will export that good.
- A country with the higher price for a good in terms of the other good and compared to the other country will import that good.
- Trade will push the lower autarky price ratio up and the higher autarky price ratio down.
- The free trade price ratio (or terms of trade) will be equal in both countries and will lie between the two countries' autarky terms of trade.
- Profit-seeking behavior in a market will induce firms to export the comparative advantage good.
- Profit-seeking behavior in a market will induce a country to specialize in the comparative advantage good.

EXERCISES

1. Identify which country exports cheese if in autarky 1 lb. of cheese trades for 2 gals. of wine in Australia and 3 gals. of wine in New Zealand.
2. Suppose Canada and Brazil are defined by a Ricardian model and have exogenous variables with the values below.

TABLE 2.12 EXOGENOUS VARIABLE VALUES

Canada	$a_{LC} = 10$	$a_{LW} = 20$	$L = 24$
Brazil	$a_{LC}^* = 5$	$a_{LW}^* = 15$	$L^* = 24$

where

L = the labor endowment in Canada (the total number of hours the workforce is willing to provide)

a_{LC} = unit labor requirement in cheese production in Canada
(hours of labor necessary to produce one unit of cheese)

a_{LW} = unit labor requirement in wine production in Canada
(hours of labor necessary to produce one unit of wine)

*All starred variables are defined in the same way but refer to the process in Brazil.

- a. Calculate the autarky terms of trade in each country.
- b. Identify the trade pattern that would arise.
- c. Specify a plausible free trade price ratio.

2.10 Welfare Effects of Free Trade: Real Wage Effects

LEARNING OBJECTIVES

1. Learn why real wages are an appropriate way to measure individual well-being.
2. Learn how the real wages formulae are derived from zero-profit conditions.

There are two ways to evaluate the welfare effects of trade in the Ricardian model. The first method evaluates the real wages of workers as two countries move from autarky to free trade. It is shown that the purchasing power of all workers' wages in both countries would rise in moving to free trade.

The focus on real wages allows us to see the effect of free trade on individual consumers in the economy. Nominal wages are not sufficient to tell us if workers gain since, even if wages rise, the price of one of the goods also rises when moving to free trade. If the price rises by a greater percentage than the wage, the ability to purchase that good falls and the worker may be worse off.

For this reason, we must consider real wages. The **real wage**¹² represents the purchasing power of wages—that is, the quantity of goods the wages will purchase. Real wages are typically measured by dividing nominal wages by a price index. The price index measures the average level of prices relative to a base year. The nominal wage is the amount of dollars the worker receives.

In this model, we need not construct a price index since there are only two goods. Instead, we will look at the real wage of workers in terms of the purchasing power of each good. In other words, we will solve for a real wage in terms of purchases of both wine and cheese.

Numerical Example: Calculating a Real Wage

12. The quantity of a good that can be purchased per unit of work. Real wage is a measure of the purchasing power of a wage and is an effective measure of well-being.

Consider the real wage of a worker in terms of cheese. Suppose the worker earns \$10 per hour and the price of cheese is \$5 per pound. The real wage can be found by dividing the wage by the price to get

$$\frac{w}{P_C} = \frac{\$10/hr}{\$5/lb} = 2\text{lbs}/hr.$$

This means the worker can buy two pounds of cheese with every hour of work.

The Real Wage of Cheese Workers in Terms of Cheese

The real wage of cheese workers in terms of cheese is the quantity of cheese that a cheese worker can buy with a unit of work. It is calculated by dividing the worker's wage by the price of cheese, written as $\frac{w_C}{P_C}$. Since zero profit results in each producing industry, we can simply rewrite the relationship derived above to construct the following formula for the real wage:

$$\frac{w_C}{P_C} = \frac{1}{a_{LC}}.$$

This means that the real wage of a worker in terms of how much cheese can be purchased is equal to labor productivity in cheese production. In other words, the amount of cheese that a worker can *buy* per period of work is exactly the same as the amount of cheese the worker can *make* in that same period.

The Real Wage of Cheese Workers in Terms of Wine

The real wage of cheese workers in terms of wine is the quantity of wine that a cheese worker can buy with a unit of work. It is calculated by dividing the cheese worker's wage by the price of wine and is written as $\frac{w_C}{P_W}$. Using the relationship between wages and prices when zero profit results in the cheese industry implies that

$$\frac{w_C}{P_W} = \frac{\left(\frac{P_C}{a_{LC}}\right)}{P_W} = \frac{1}{a_{LC}} \frac{P_C}{P_W}.$$

This means that the real wage of cheese workers in terms of wine is the product of labor productivity in the cheese industry and the price ratio. Labor productivity gives the quantity of cheese a cheese worker makes in an hour of work. The price ratio gives the quantity of wine that exchanges for each unit of cheese. The product gives the quantity of wine that a cheese worker can buy with a unit of work. To calculate the autarky real wage, simply plug in the autarky price ratio. To calculate the free trade real wage, plug in the free trade price ratio.

The Real Wage of Wine Workers in Terms of Wine

The real wage of wine workers in terms of wine is the quantity of wine that a wine worker can buy with a unit of work. It is calculated by dividing the worker's wage by the price of wine, written as w_W/P_W . Since zero profit results in each producing industry, we can rewrite the relationship to get

$$\frac{w_W}{P_W} = \frac{1}{a_{LW}}.$$

As with cheese, the real wage of a worker in terms of how much wine can be purchased is equal to labor productivity in wine production. In other words, the amount of wine that a worker can buy per period of work is exactly the same as the amount of wine the worker can make in that same period.

The Real Wage of Wine Workers in Terms of Cheese

The real wage of wine workers in terms of cheese is the quantity of cheese that a wine worker can buy with a unit of work. It is calculated by dividing the wine worker's wage by the price of cheese, written as (w_W/P_C) . Using the relationship between prices and wages when zero profit results in the wine industry implies that

$$\frac{w_W}{P_C} = \frac{\left(\frac{P_W}{a_{LW}}\right)}{P_C} = \frac{1}{a_{LW}} \frac{P_W}{P_C}.$$

This means that the real wage of wine workers in terms of cheese is the product of labor productivity in the wine industry and the price ratio. Labor productivity gives the quantity of wine a wine worker makes in an hour of work. The price ratio gives the quantity of cheese that exchanges for each unit of wine. The product gives the quantity of cheese that a wine worker can buy with a unit of work. To solve for the autarky real wage, simply plug in the autarky price ratio. To find the free trade real wage, plug in the free trade price ratio.

Real Wages in Autarky

To calculate autarky real wages, we simply plug the autarky price ratio into the real wage formulae.

Recall that the autarky price ratio is $\left(\frac{P_C}{P_W}\right)^{Aut} = \frac{a_{LC}}{a_{LW}}$. Plugging this in and simplifying yields the results in Table 2.13 "Autarky Real Wages".

Table 2.13 Autarky Real Wages

	In Terms of Cheese	In Terms of Wine
Real Wage of Cheese Workers	$\frac{w_C}{P_C} = \frac{1}{a_{LC}}$	$\frac{w_C}{P_W} = \frac{1}{a_{LC}} \frac{a_{LC}}{a_{LW}} = \frac{1}{a_{LW}}$
Real Wage of Wine Workers	$\frac{w_W}{P_C} = \frac{1}{a_{LW}} \frac{a_{LW}}{a_{LC}} = \frac{1}{a_{LC}}$	$\frac{w_W}{P_W} = \frac{1}{a_{LW}}$

where

P_C = price of cheese

P_W = price of wine

w_C = wage paid to cheese workers

w_W = wage paid to wine workers

a_{LC} = unit labor requirement in cheese production in the United States (hours of labor necessary to produce one unit of cheese)

a_{LW} = unit labor requirement in wine production in the United States (hours of labor necessary to produce one unit of wine)

Notice that in autarky, the real wage of cheese workers is exactly the same as the real wage of wine workers with respect to purchases of both goods. This occurs because labor is assumed to be homogeneous—that is, all labor is the same—and because there is free mobility between industries. (If workers were paid different wages, the lower-wage workers would move to the higher-wage industry.)

Comparison of Autarky Real Wages between Countries

Suppose the United States has an absolute advantage in the production of both goods. In this case, $\frac{1}{a_{LC}} > \frac{1}{a_{LC}^*}$ and $\frac{1}{a_{LW}} > \frac{1}{a_{LW}^*}$. This implies that the real wages of workers in both industries in the United States are higher than the real wages in France. Put another way, workers in France earn lower wages in both industries.

Sometimes cross-country wage comparisons are made and it is suggested that firms in a high-wage country cannot compete with firms in low-wage countries. However, wage comparisons of this kind are not sufficient in this model to determine who will produce what or whether trade can be advantageous. Instead, what matters is relative wage comparisons. In this model, a country will tend to specialize in the good in which it has the greatest real wage advantage. Thus if

$$\frac{\frac{1}{a_{LC}}}{\frac{1}{a_{LC}^*}} > \frac{\frac{1}{a_{LW}}}{\frac{1}{a_{LW}^*}},$$

then the United States has relatively higher real wages with respect to cheese purchases than it does in wine purchases. When trade opens, the United States will specialize in its comparative advantage good, which, by rearranging the above inequality, can easily be shown to be cheese.

Effects of Free Trade on Real Wages

Suppose two countries, the United States and France, move from autarky to free trade. If the United States has the comparative advantage in cheese production, then $\frac{a_{LC}}{a_{LW}} < \frac{a_{LC}^*}{a_{LW}^*}$, which implies $\left(\frac{P_C}{P_W}\right)^{Aut} < \left(\frac{P_C^*}{P_W^*}\right)^{Aut}$. When the two countries move to free trade, the free trade price ratio will lie somewhere between the autarky price ratios. This means that (P_C/P_W) rises in the United States when moving from autarky to free trade, while P_C^*/P_W^* falls when moving to free trade.

The other major change that occurs is that the United States specializes in cheese production, while France specializes in wine production. This means that real wages in free trade for wine workers in the United States need not be calculated since the United States will no longer have any wine workers. Similarly, real wages for cheese workers in France need not be calculated.

Thus we can calculate the changes in real wages shown in [Table 2.14 "Changes in Real Wages \(Autarky to Free Trade\)"](#).

Table 2.14 Changes in Real Wages (Autarky to Free Trade)

	In Terms of Cheese	In Terms of Wine
Real Wage of U.S. Cheese Workers	$\frac{w_C}{P_C} = \frac{1}{a_{LC}}$ (no change)	$\frac{w_C}{P_W} = \frac{1}{a_{LC}} \frac{P_C}{P_W}$ (rises)
Real Wage of French Wine Workers	$\frac{w_W}{P_C} = \frac{1}{a_{LW}} \frac{P_W}{P_C}$ (rises)	$\frac{w_W}{P_W} = \frac{1}{a_{LW}}$ (no change)

First, consider the fate of U.S. cheese workers. Since the unit labor requirement for cheese does not change in moving to free trade, there is also no change in the real wage in terms of cheese. However, since the price of cheese in terms of wine rises, U.S. cheese workers can get more wine for each unit of cheese in exchange. Thus the real wage of cheese workers in terms of wine rises. This means cheese workers are at least as well off in free trade as they were in autarky.

The worst outcome occurs if a cheese worker has no demand for wine. Perhaps an individual abstains from alcohol consumption. In this case, the worker would be able to buy just as much cheese in free trade as in autarky, but no more. Such a person would receive no benefit from free trade. However, every worker who demands both wine and cheese will be able to buy more of both goods.

As for the workers who worked in the wine industry in the United States in autarky, they are now cheesemakers earning cheesemaker wages. Since real wages for wine workers were the same as wages for cheese workers in autarky, and since cheese workers are no worse off with free trade, then wine workers must also be no worse off in free trade. Of course, the model assumes that the movement of workers from one industry to another is costless. In the immobile factor model, we address the implications of adjustment costs across industries.

In France, the real wage of winemakers in terms of how much wine they can buy remains constant, while the real wage in terms of cheese must go up. French cheesemakers have all become winemakers because of specialization, which means all French workers are no worse off and most likely better off as a result of free trade.

The likely welfare effect of free trade, then, is that everyone in both trading countries benefits. At the very worst, some individuals will be just as well off as in autarky. This result occurs for any free trade price ratio that lies between the autarky price ratios.

In David Ricardo's original numerical example, he demonstrated that when both countries specialize in their comparative advantage goods and engage in free trade, both countries can experience gains from trade. However, his demonstration was only true for particular numerical values. By calculating real wage changes, it is shown that it doesn't matter which price ratio emerges in free trade as long as it is between the autarky prices. Also, because all workers receive the same wage in each country, the real wage calculations tell us that everyone benefits equally in each country.

KEY TAKEAWAYS

- Real wages are an appropriate measure of worker well-being because they represent the purchasing power of the wage.
- Real wages are positively related to labor productivity in the Ricardian model.
- When countries move to free trade, the real wage with respect to the exported good remains constant, but the real wage with respect to the imported good rises in both countries.
- If workers prefer to consume a positive amount of both goods, then when a country moves to free trade, every worker will be able to buy more of both goods. In other words, everyone in both countries will benefit from trade.

EXERCISE

1. Consider a Ricardian model. Suppose the U.S. unit labor requirement for timber is three, its unit labor requirement for videocassette recorders (VCRs) is eight, and it has forty-eight million workers. Suppose Taiwan's unit labor requirement for timber is six, its unit labor requirement for VCRs is two, and it has forty-eight million workers.
 - a. Which country has the absolute advantage in each good? Which country has the comparative advantage? Explain.
 - b. Calculate each country's autarky price ratio. Then make up a plausible free trade price ratio. What are the levels of production and the pattern of trade when free trade occurs?
 - c. Calculate real wages for workers in both countries in autarky and free trade. Explain why everyone benefits from trade.
 - d. Suppose the United States implements a costless technology improvement program that lowers the U.S. unit labor requirement for timber to two. What effect would this have on the world supply of timber? What effect would this have on the free trade price ratio? Explain how real wages would change in both the United States and Taiwan.

2.11 The Welfare Effects of Free Trade: Aggregate Effects

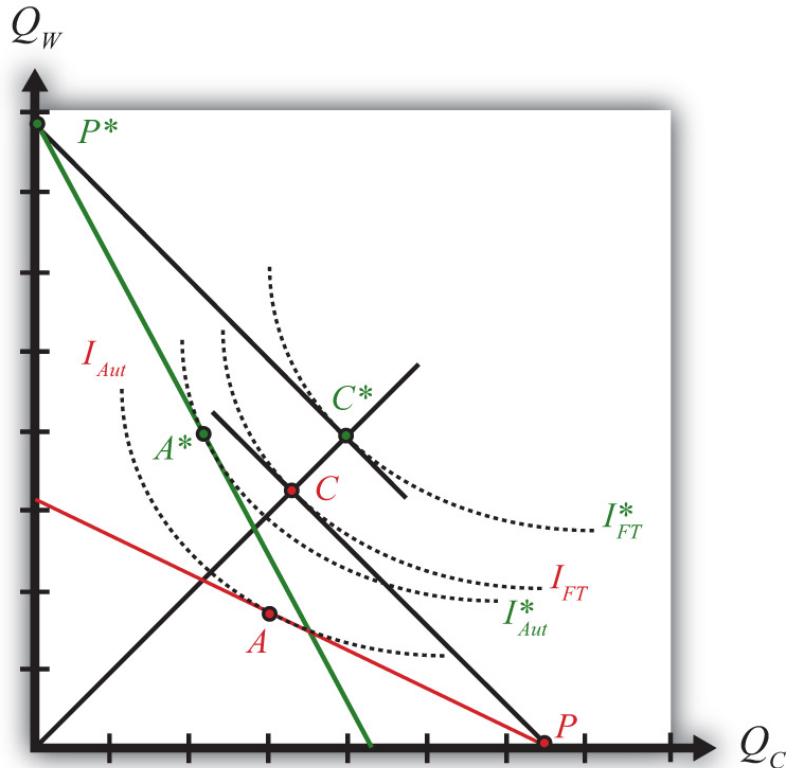
LEARNING OBJECTIVE

1. Learn how national welfare can rise for both countries when moving to free trade in a Ricardian model.

The second and more traditional method to evaluate the effects of free trade uses an aggregate welfare function to depict the overall welfare effects that would accrue to the nation. This method allows one to demonstrate the benefits that arise from increased production and consumption efficiency.

Figure 2.5 "Comparing Autarky and Free Trade Equilibriums" compares autarky and free trade equilibriums for the United States and France. The U.S. PPF is given by the red line, while France's PPF is given by the green line. We assume both countries share the same aggregate preferences represented by the indifference curves in the diagram. Note also that if the United States and France had the same size labor force, then the relative positions of the PPFs imply that the United States has the absolute advantage in cheese production, while France has the absolute advantage in wine production. Also, if each country has an absolute advantage in one of the two goods, then each country must also have the comparative advantage in that good.

Figure 2.5 Comparing Autarky and Free Trade Equilibriums



The U.S. autarky production and consumption points are determined where the aggregate indifference curve is tangent to the U.S. PPF. This occurs at the red point A. The United States realizes a level of aggregate utility that corresponds to the indifference curve I_{Aut} .

The U.S. production and consumption points in free trade are at the red P and C, respectively. The United States specializes in production of its comparative advantage good but trades to achieve its consumption point at the red C. In free trade, the United States realizes a level of aggregate utility that corresponds to the indifference curve I_{FT} . Since the free trade indifference curve I_{FT} lies to the northeast of the autarky indifference curve I_{Aut} , national welfare rises as the United States moves to free trade.

France's autarky production and consumption points are determined by finding the aggregate indifference curve that is tangent to the French PPF. This occurs at the green point A^* . France realizes a level of aggregate utility that corresponds to the indifference curve I_{Aut}^* .

French production and consumption points in free trade are the green P^* and C^* , respectively. In free trade, France realizes a level of aggregate utility that corresponds to the indifference curve I_{FT^*} . Since the free trade indifference curve I_{FT^*} lies to the northeast of the autarky indifference curve I_{Aut^*} , national welfare rises as France moves to free trade.

KEY TAKEAWAYS

- National welfare can be represented with a set of aggregate indifference curves plotted in a PPF diagram.
- Free trade will raise aggregate welfare for both countries relative to autarky. Both countries are better off with free trade.

EXERCISE

1. Suppose each country specialized in the wrong good. Depict an equilibrium using the free trade prices in each country to show why national welfare would fall in free trade relative to autarky.

2.12 Appendix: Robert Torrens on Comparative Advantage

The first known statement of the principle of comparative advantage and trade appears in an article by Robert Torrens in 1815 titled *Essay on the External Corn Trade*. Torrens begins by describing the basic idea of absolute advantage as described by Adam Smith but goes on to suggest that the simple intuition is erroneous. He wrote,

Suppose that there are in England, unreclaimed districts, from which corn might be raised at as small an expense of labor and capital, as from the fertile plains of Poland. This being the case, and all other things the same, the person who should cultivate our unreclaimed districts, could afford to sell his produce at as cheap a rate as the cultivator of Poland: and it seems natural to conclude, that if industry were left to take its most profitable direction, capital would be employed in raising corn at home, rather than bringing it in from Poland at an equal prime cost, and at much greater expense of carriage. But this conclusion, however obvious and natural it may, at first sight, appear, might, on closer examination, be found entirely erroneous. If England should have acquired such a degree of skill in manufactures, that, with any given portion of her capital, she could prepare a quantity of cloth, for which the Polish cultivator would give a greater quantity of corn, then she could, with the same portion of capital, raise from her own soil, then, tracts of her territory, though they should be equal, nay, even though they should be superior, to the lands in Poland, will be neglected; and a part of her supply of corn will be imported from that country.

In the first part of the passage, Torrens considers a case in which the cost of producing corn, in terms of labor and capital usage, is the same in England as it is in Poland. He points out that producers could afford to sell both English and Polish corn at the same low price. However, since it would cost additional resources to transport the corn from Poland to England (expense of carriage), it makes intuitive sense that corn should be produced in England, rather than imported, since Polish corn would wind up with a higher price than English corn in the English market.

He continues by suggesting that this conclusion is erroneous. Why? Suppose England were to remove some capital (and labor) from the production of corn and move it into the production of manufactured goods. Suppose further that England trades this newly produced quantity of manufactured goods for corn with Poland. This outcome would be better for England if the amount of corn that Poland is willing to trade for the manufactured goods is greater than the amount of corn that England has given up producing. If the excess corn that Poland is willing to trade is sufficiently large, then it may be more than enough to pay for the transportation costs between the two countries. Torrens's final point is that this trading outcome

may be superior for England even if the lands of England should be superior to the lands of Poland—in other words, even if corn can be more efficiently produced in England (i.e., at lower cost) than in Poland.

This is the first explicit description of one of the major results from the theory of comparative advantage. It reflects Torrens's understanding that a country might conceivably benefit from free trade while reducing or eliminating production of a good it is technologically superior at producing.

Chapter 3

The Pure Exchange Model of Trade

The pure exchange model is one of the most basic models of trade and is even simpler than the Ricardian model in Chapter 2. The model develops a simple story: What if one person who possesses one type of good (say apples) meets up with another person who possesses another type of good (say oranges)? What could we say about two people trading apples for oranges?

As it turns out, we can say quite a bit. The pure exchange model demonstrates the advantages of mutually voluntary exchange. And when the simple story is extended to include a second apple seller, the model shows the positive and negative effects associated with competition. When the competition is from another country, the model demonstrates how international trade can generate both winners and losers in the economy. This chapter offers the first example showing that trade can cause a redistribution of income, with some winning from trade and others losing from trade.

3.1 A Simple Pure Exchange Economy

LEARNING OBJECTIVES

1. Learn the definition of the terms of trade.
2. Learn how the terms of trade between two goods is equivalent to the ratio of dollar prices for the two goods.

The Ricardian model shows that trade can be advantageous for countries. If we inquire deeper and ask what is meant when we say a “country” benefits in this model, we learn it means that every individual, every worker, in both countries is able to consume more goods after specialization and trade. In other words, everyone benefits from trade in the Ricardian model. Everybody wins.

Unfortunately, though, this outcome is dependent on the assumptions made in the model, and in some important ways these assumptions are extreme simplifications. One critical assumption is that the workers in each country are identical; another is the free and costless ability of workers to move from one industry to another. If we relax or change these assumptions, the win-win results may not remain. That's what we will show in the pure exchange model and the immobile factor model.

For a variety of reasons, it is more common for trade to generate both winners and losers instead of all winners. Economists generally refer to a result in which there are both winners and losers as **income redistribution**¹ because the winners can be characterized as receiving a higher real income, while those who lose suffer from a lower real income.

The simplest example of advantageous trade arising from differences in resource endowments can be shown with a pure exchange model. In this model, we ignore the production process and assume more simply that individuals are endowed with a stock of consumption goods. We also show that trade can result in a redistribution of income. The model and story are adapted from a presentation by James Buchanan about the benefits of international trade. James Buchanan, “The Simple Logic of Free Trade,” *Proceedings of the First Annual Symposium of the Institute for International Competitiveness* (Radford, VA: Radford University, 1988), iii–x.

1. Occurs when some individuals gain income while others lose or when individuals gain and lose income shares of total income.

A Simple Example of Trade

Suppose there are two individuals: Farmer Smith and Farmer Jones. Farmer Smith lives in an orange grove, while Farmer Jones lives in an apple orchard. For years, these two farmers have sustained themselves and their families by collecting oranges and apples on their properties: Smith eats only oranges and Jones eats only apples.

One day these two farmers go out for a walk. Farmer Smith carries ten oranges with him in case he becomes hungry. Farmer Jones carries ten apples. Suppose these farmers meet. After a short conversation, they discover that the other farmer sustains his family with a different product, and the farmers begin to discuss the possibility of a trade.

The farmers consider trade for the simple reason that each prefers to consume a variety of goods. We can probably imagine the monotony of having to eat only apples or only oranges day after day. We can also probably imagine that having both apples and oranges would be better, although we might also prefer some fried chicken, mashed potatoes, a Caesar salad, and numerous other favorite foods, but that is not included as a choice for these farmers. As such, when we imagine trade taking place, we are also assuming that each farmer has a preference for variety in consumption. In some special cases, this assumption may not be true. For example, Farmer Jones might have a distaste for oranges, or he may be allergic to them. In that special case, trade would not occur.

Assuming trade is considered by the farmers, one question worth asking is, What factors will determine the **terms of trade**²? The terms of trade is defined as the quantity of one good that exchanges for a quantity of another. In this case, how many apples can be exchanged for how many oranges? It is typical to express the terms of trade as a ratio. Thus, if one apple can be exchanged for four oranges, we can write the terms of trade as follows:

$$\text{TOT} = \frac{1 \text{ apple}}{4 \text{ oranges}} = \frac{1}{4} \text{ apple/orange},$$

where TOT refers to terms of trade. It is immaterial whether the ratio is written apples over oranges or oranges over apples, but to proceed, one or the other must be chosen.

2. The amount of one good traded per unit of another in a mutually voluntary exchange. Often expressed as a ratio of prices.

The terms of trade is also equivalent to the ratio of prices between two goods. Suppose P_A is the price of apples (measured in dollars per apple) and P_O is the price of oranges (measured in dollars per orange). Then

$$\text{TOT} = \frac{P_O}{P_A} \left[\frac{\frac{\$}{\text{orange}}}{\frac{\$}{\text{apple}}} = \frac{\$}{\text{orange}} \times \frac{\text{apple}}{\$} = \frac{\text{apples}}{\text{orange}} \right].$$

To demonstrate the equivalency, consider the units of this price ratio shown in brackets above. After some manipulation, we can see that the dollars cancel and thus the price of oranges over the price of apples is measured in units of apples per orange. We can refer to this price ratio as the price of oranges in terms of apples—that is, how many apples one can get in exchange for every orange. Notice that the price of oranges over apples is in units of *apples* per orange. Similarly, P_A/P_O has units of oranges per apple. This model and many others we will consider are actually barter economies. This means that no money is being exchanged between the agents. Instead, one good is exchanged for another good. However, since we are accustomed to evaluating values in monetary terms, we will often write important expressions, like the terms of trade, in terms of their monetary equivalents as we have done here.

KEY TAKEAWAYS

- The terms of trade is defined as how much of one good trades for one unit of another good in the market.
- The terms of trade between two goods (e.g., apples and oranges) is equivalent to the ratio of the dollar prices of apples and oranges.

EXERCISES

1. If two bushels of apples can be traded for three bushels of oranges, what is the terms of trade between apples and oranges?
2. If two bushels of apples can be traded for three bushels of oranges, how many bushels of oranges can be purchased with one bushel of apples?
3. If the price of ice cream is \$3.50 per quart and the price of cheesecake is \$4.50 per slice, what is the terms of trade between cheesecake and ice cream?

3.2 Determinants of the Terms of Trade

LEARNING OBJECTIVES

1. Understand how the terms of trade for any two products between any two people will be affected by a wide variety of factors.
2. Recognize that many of the determinants correspond to well-known concerns in business and ethics.

The terms of trade ultimately decided on by the two trading farmers will depend on a variety of different and distinct factors. Next we describe many of these factors.

Preferences

The strength of each farmer's desire for the other product will influence how much he is willing to give up to obtain the other product. Economists assume that most products exhibit diminishing marginal utility. This means that the tenth orange consumed by Farmer Smith adds less utility than the first orange he consumes. In effect, we expect people to get tired of eating too many oranges. Since for most people the tenth orange consumed will be worth less than the first apple consumed, Farmer Smith would be willing to trade at least one orange for one apple. As long as the same assumption holds for Farmer Jones, the tenth apple for him will be worth less than the first orange, and he will be willing to trade at least one for one. How many more oranges might trade for how many more apples will depend on how much utility each farmer gets from successive units of both products: in other words, it depends on the farmers' preferences.

Uncertainty

In this situation, each farmer is unlikely to have well-defined preferences. Farmer Smith may never have tasted an apple, and Farmer Jones may never have tasted an orange. One simple way to resolve this uncertainty is for the farmers to offer free samples of their products before an exchange is agreed on. Without a sample, the farmers would have to base their exchanges on their expectations of how they will enjoy the other product. Free samples, on the other hand, can be risky. Suppose a sample of oranges is provided and Farmer Jones learns that he hates the taste of oranges. He might decide not to trade at all.

To overcome uncertainty in individual preferences, many consumer products are offered in sample sizes to help some consumers recognize that they do have a preference for the product. This is why many supermarkets offer free samples in their aisles and why drink companies sometimes give away free bottles of their products.

Scarcity

The relative quantities of the two goods available for trade will affect the terms of trade. If Farmer Smith came to the market with one hundred oranges to Farmer Jones's ten apples, then the terms of trade would likely be different than if the farmers came to the market with an equal number. Similarly, if the farmers came to the market with ten oranges and ten apples, respectively, but recognized that they had an entire orchard of apples and an entire grove of oranges waiting back at home, then the farmers would be more likely to give up a larger amount of their product in exchange.

Size

The sizes of the apples and oranges are likely to influence the terms of trade. One would certainly expect that Farmer Smith would get more apples for each orange if the oranges were the size of grapefruits and the apples the size of golf balls than if the reverse were true.

Quality

The quality of the fruits will influence the terms of trade. Suppose the apples are sweet and the oranges are sour. Suppose the apples are filled with worm holes. Suppose the oranges are green rather than orange. Or consider the vitamin, mineral, and calorie contents of each of the fruits. Quality could also be assessed by the variety of uses for each product. For example, apples can be eaten raw, turned into applesauce, squeezed into juice, made into pies, or covered with caramel.

Effort

Although a pure exchange model assumes that no production takes place, imagine momentarily that some effort is required to harvest the fruit. What if apples grew at the top of tall trees that required a precarious climb? What if predatory wolves lived in the orange grove? Surely these farmers would want to take these factors into account when deciding the terms for exchange. Of course, this factor is related to scarcity. The more difficult it is to produce something, the scarcer that item will be.

Persuasion

The art of persuasion can play an important role in determining the terms of trade. Each farmer has an incentive to embellish the quality and goodness of his product and perhaps diminish the perception of quality of the other product. Farmer Smith might emphasize the high quantities of vitamin C found in oranges while noting that apples are relatively vitamin deficient. He might argue that oranges are consumed by beautiful movie stars who drive fast cars, while apples are the food of peasants. He might also underemphasize his own desire for apples. The more persuasive Farmer Smith is, the more likely he is to get a better deal in exchange. Note that the farmer's statements need not be truthful as long as the other farmer is uncertain about the quality of the other product. In this case, differences in the persuasive abilities of the two farmers can affect the final terms of trade.

Expectations of Utility

Decisions about how much to trade are based on the utility one expects to obtain upon consuming the good. The utility one ultimately receives may be less. Indeed, in some cases the value of what one receives may be less than the value of what one gives up. However, this outcome will arise only if expectations are not realized.

For example, a person may choose to voluntarily pay \$10 to see a movie that has just been released. Perhaps the person has read some reviews of the movie or has heard from friends that the movie is very good. Based on prior evaluation, the person decides that the movie is worth at least \$10. However, suppose this person winds up hating the movie and feels like it was a complete waste of time. In hindsight, with perfect knowledge about his own preferences for the movie, he might believe it is only worth \$5 or maybe just \$2, in which case he is clearly worse off after having paid \$10 to see the movie. This is one reason individuals may lose from trade, but it can only occur if information is imperfect.

Expectations of a Future Relationship

If the farmers expect that the current transaction will not be repeated in the future, then there is a potential for the farmers to misrepresent their products to each other. Persuasion may take the form of outright lies if the farmers do not expect to meet again. Consider the traveling medicine man portrayed in U.S. Western movies. He passes through town with a variety of elixirs and promises that each will surely cure your ailment and possibly do much more. Of course, chances are good that the elixirs are little more than colored water with some alcohol and are unlikely to cure anything. But this type of con game is more likely when only one transaction is expected. However, if the transaction is hoped to be the first of many to come, then untruthful embellishments will be less likely.

Government Policies

If a taxman stands ready to collect a tax based on the amounts traded between the two farmers, this is likely to affect the terms of trade. Also, if laws impose penalties for misrepresentation of a product, then this will also affect the farmers' behavior in determining the terms of trade.

Morality

Imagine that Farmer Smith was raised to always tell the truth, while Farmer Jones missed those lessons during his upbringing. In this case, Farmer Jones might be more likely to misrepresent his apples in order to extract a more favorable terms of trade.

Coercion

Finally, the terms of trade can also be affected by coercion. If Farmer Jones threatens Farmer Smith with bodily injury, he might be able to force an exchange that Farmer Smith would never agree to voluntarily. At the extreme, he could demand all of Farmer Smith's oranges and not give up any apples in exchange. Of course, once coercion enters a transaction, it may no longer be valid to call it trade—it would be more accurate to call it theft.

Summary

Notice that many of these determinants relate to good business practices and ethical behavior. Business schools have classes in marketing and product promotion, sales advertising, and quality control, all of which can be thought of as ways to improve the terms of trade for the product the business is selling. Ethics teaches one to be truthful and to represent one's products honestly. It also teaches one not to steal or use force to obtain what one desires.

How all these factors play into the matter ultimately influences what the terms of trade will be between products. As such, this simple model of trade can be embellished into a fairly complex model of trade. That some terms of trade will arise is simple to explain. But what precisely will be the terms of trade involves a complex mixture of factors.

KEY TAKEAWAY

- The terms of trade is influenced by many different factors, including product preferences, uncertainties over preferences, quantities and qualities of the goods, persuasive capabilities, regularity of the trading relationship, and government policies.

EXERCISES

1. Give an example, from your own experience perhaps, in which the expected benefits from trade are positive but the actual benefits from trade are negative.
2. Suppose Larry initially proposes to give Naomi twenty music CDs in exchange for a ride to Atlanta. How would the final terms of trade change if each of the following occurs before the deal is settled?
 - a. Larry learns that Naomi's car has no air conditioning and the temperature that day will be ninety-five degrees.
 - b. Naomi tells Larry that her beautiful cousin may travel with them.
 - c. Naomi mentions that none of the CDs are by her favorite artists.
 - d. Larry learns that Naomi will also be bringing her two dogs and three cats.
 - e. Naomi tells Larry that she will be able to borrow her Dad's 600 series BMW.
 - f. Larry hopes to be able to get rides from Naomi in the future too.

3.3 Example of a Trade Pattern

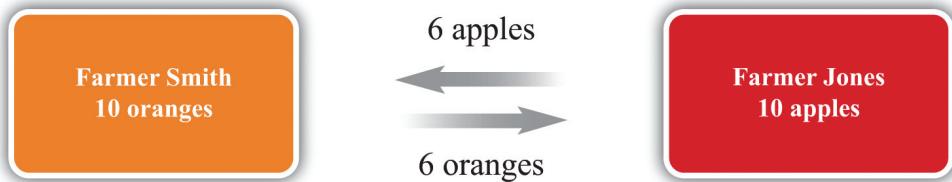
LEARNING OBJECTIVE

1. Learn how to describe a mutually voluntary exchange pattern and specify both the terms of trade and the final consumption bundles for two traders.

Suppose after some discussion Farmer Smith and Farmer Jones agree to a **mutually voluntary exchange**³ of six apples for six oranges (see [Figure 3.1 "Two-Farmer Trade Pattern"](#)). The terms of trade is six apples per six oranges, or one apple per orange. After trade, Farmer Smith will have four oranges and six apples to consume, while Farmer Jones will have six oranges and four apples to consume. As long as the trade is voluntary, it must hold that both farmers expect to be better off after trade since they are free not to trade. Thus mutually voluntary trade must be beneficial for both farmers.

Figure 3.1 Two-Farmer Trade Pattern

Trade Pattern



$$\text{TOT} = P_O/P_A = 6 \text{ apples}/6 \text{ oranges} = 1 \text{ apple/orange}$$

Posttrade Consumption



3. A trade of one item for another chosen willingly (i.e., without coercion) by both individuals in a market.

Sometimes people talk about trade as if it were adversarial, with one side competing against the other. With this impression, one might believe that trade

would generate a winner and a loser as if trade were a contest. However, a pure exchange model demonstrates that trade is not a zero-sum game. Instead, when two individuals make a voluntary exchange, they will both benefit. This is sometimes called a positive-sum game. A zero-sum game is a contest whose outcome involves gains and losses of equal value so that the sum of the gains and losses is zero. In contrast, a positive-sum game is one whose outcome involves total gains that exceed the total losses so that the sum of the gains and losses is positive.

Sometimes the pure exchange model is placed in the context of two trading countries. Suppose instead of Farmer Smith and Farmer Jones, we imagine the United States and Canada as the two “individuals” who trade with each other. Or, better still, we might recognize that international trade between countries consists of millions, or billions, of individual trades much like the one described here. If each individual trade is mutually advantageous, then the summation of billions of such trades must also be mutually advantageous. Thus, as long as the people within each country can choose not to trade if they so desire, trade must be beneficial for every trader in both countries.

Nonetheless, although this conclusion is sound, it is incorrect to assert that everyone in each country will necessarily benefit from free trade. Although the national effects will be positive, a country is composed of many individuals, many of whom do not engage in international trade. Trade can make some of them worse off. In other words, trade is likely to cause a redistribution of income, generating both winners and losers. This outcome is first shown in [Chapter 3 "The Pure Exchange Model of Trade"](#), [Section 3.4 "Three Traders and Redistribution with Trade"](#).

KEY TAKEAWAYS

- Any trade pattern between individuals may be claimed to be mutually advantageous as long as the trade is mutually voluntary.
- The terms of trade is defined as the ratio of the trade quantities of the two goods.
- The final consumption bundles are found by subtracting what one gives away and adding what one receives to one’s original endowment.

EXERCISE

1. Suppose Kendra has ten pints of milk and five cookies and Thomas has fifty cookies and one pint of milk.
 - a. Specify a plausible mutually advantageous trading pattern.
 - b. Identify the terms of trade in your example (use units of pints per cookie).
 - c. Identify the final consumption bundles for Kendra and Thomas.
 - d. Which assumption or assumptions guarantee that the final consumption bundles provide greater utility than the initial endowments for both Kendra and Thomas?

3.4 Three Traders and Redistribution with Trade

LEARNING OBJECTIVES

1. Learn how changes in the numbers of traders changes the terms of trade and affects the final consumption possibilities.
2. Learn that an increase in competition causes a redistribution of income.
3. Learn the importance of the profit-seeking assumption to the outcome.
4. Learn how one's role as a seller or buyer in a market, affects one's preference for competition.

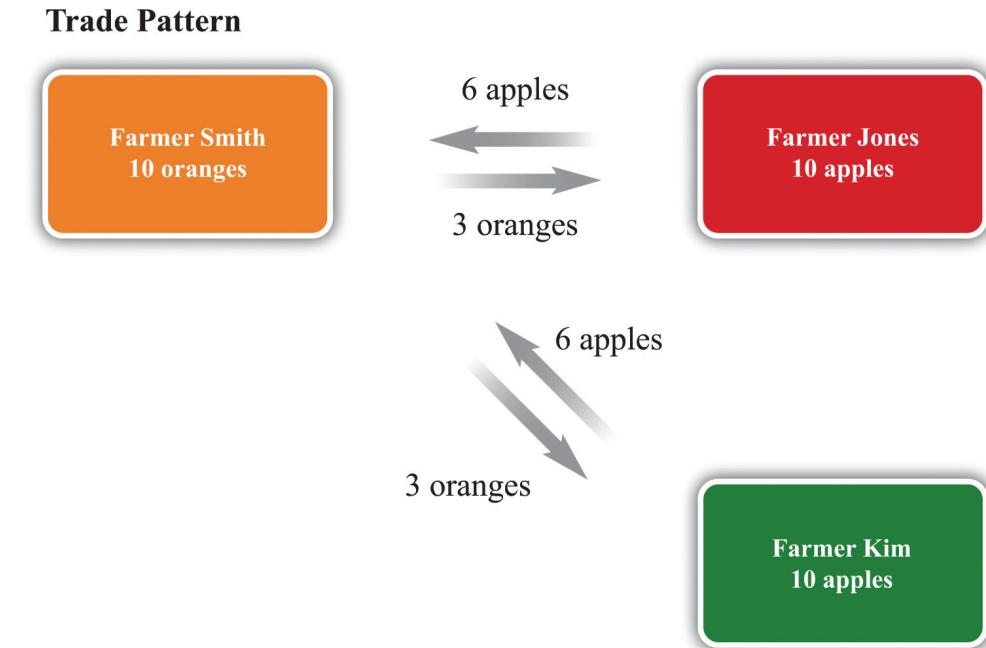
Suppose for many days, months, or years, Farmer Smith and Farmer Jones are the only participants in the market. However, to illustrate the potential for winners and losers from trade, let us extend the pure exchange model to include three farmers rather than two. Suppose that one day a third farmer arrives at the market where Farmer Jones and Farmer Smith conduct their trade. The third farmer is Farmer Kim, and he arrives at the market with an endowment of ten apples.

The main effect of Farmer Kim's arrival is to change the relative scarcity of apples to oranges. On this day, the total number of apples available for sale has risen from ten to twenty. Thus apples are relatively more abundant, while oranges are relatively scarcer. The change in relative scarcities will undoubtedly affect the terms of trade that is decided on during this second day of trading.

Farmer Smith, as a seller of oranges (the relatively scarcer good), now has a stronger negotiating position than he had on the previous day. Farmer Jones and Farmer Kim, as sellers of apples, are now competing against each other. With the increased supply of apples at the market, the price of apples in exchange for oranges can be expected to fall. Likewise, the price of oranges in exchange for apples is likely to rise. This means that Farmer Smith can negotiate exchanges that yield more apples for each orange compared with the previous day.

Suppose Farmer Smith negotiates a trade of three oranges for six apples with each of the two apple sellers (see [Figure 3.2 "Three-Farmer Trade Pattern"](#)). After trade, Farmer Smith will have twelve apples and four oranges for consumption. Farmers Jones and Kim will each have three oranges and four apples to consume.

Figure 3.2 Three-Farmer Trade Pattern



$$TOT = P_O/P_A = 12 \text{ apples}/6 \text{ oranges} = 2 \text{ apples/orange}$$

As before, assuming that all three farmers entered into these trades voluntarily, it must hold that each one is better off than he would be in the absence of trade. However, we can also compare the fate of each farmer relative to the previous week. Farmer Smith is a clear winner. He can now consume twice as many apples and the same number of oranges as in the previous week. Farmer Jones, on the other hand, loses due to the arrival of Farmer Kim. He now consumes fewer oranges and the same number of apples as in the previous week. As for Farmer Kim, presumably he made no earlier trades. Since he was free to engage in trade during the second week, and he agreed to do so, he must be better off.

It is worth noting that we assume here that each of the farmers, but especially Farmer Smith, is motivated by profit. Farmer Smith uses his bargaining ability because he knows that by doing so he can get a better deal and, ultimately, more goods to consume. Suppose for a moment, however, that Farmer Smith is *not* motivated by profit but instead cares about friendship. Because he and Farmer Jones had been the only traders in a market for a long period of time before the arrival of Farmer Kim, surely they got to know each other well. When Farmer Kim arrives, it is conceivable Smith will recognize that by pursuing profit, his friend Farmer Jones will lose out. In the name of friendship, Smith might refuse to trade with Kim and continue to trade at the original terms of trade with Jones. In this

case, the outcome is different because we have changed the assumptions. The trade that does occur remains mutually voluntary and both traders are better off than they were with no trade. Indeed, Smith is better off than he would be trading with Jones and Kim; he must value friendship more than more goods or else he wouldn't have voluntarily chosen this. The sole loser from this arrangement is Farmer Kim, who doesn't get to enjoy the benefits of trade.

Going back to the assumption of profit seeking, however, the example demonstrates a number of important principles. The first point is that free and open competition is not necessarily in the interests of everyone. The arrival of Farmer Kim in the market generates benefits for one of the original traders and losses for the other. We can characterize the winners and losers more generally by noting that each farmer has two roles in the market. Each is a *seller* of one product and a *buyer* of another. Farmer Smith is a seller of oranges but a buyer of apples. Farmer Jones and Farmer Kim are sellers of apples but buyers of oranges.

Farmer Kim's entrance into the market represents an addition to the number of sellers of apples and the number of buyers of oranges. First, consider Farmer Jones's perspective as a seller of apples. When an additional seller of apples enters the market, Farmer Jones is made worse off. Thus, in a free market, sellers of products are worse off the larger the number of other sellers of similar products. Open competition is simply not in the best interests of the sellers of products. At the extreme, the most preferred position of a seller is to have the market to himself—that is, to have a **monopoly**⁴ position in the market. Monopoly profits are higher than could ever be obtained in a duopoly, in an oligopoly, or with perfect competition.

Next, consider Farmer Smith's perspective as a buyer of apples. When Farmer Kim enters the market, Farmer Smith has more sources of apples than he had previously. This results in a decrease in the price he must pay and makes him better off. Extrapolating, buyers of a product will prefer to have as many sellers of the products they buy as possible. The very worst position for a buyer is to have a single monopolistic supplier. The best position is to face a perfectly competitive market with lots of individual sellers, where competition may generate lower prices.

Alternatively, consider Farmer Jones's position as a buyer of oranges. When Farmer Kim enters the market there is an additional buyer. The presence of more buyers makes every original buyer worse off. Thus we can conclude that buyers of products would prefer to have as few other buyers as possible. The best position for a buyer is a **monopsony**⁵—a situation in which he is the single buyer of a product.

4. An individual or firm that is the sole seller of a product in a market.

5. An individual or firm that is the sole buyer of a product in a market.

Finally, consider Farmer Smith's role as a seller of oranges. When an additional buyer enters the market, Farmer Smith becomes better off. Thus sellers of products would like to have as many buyers for their product as possible.

More generally, we can conclude that producers of products (sellers) should have little interest in free and open competition in their market, preferring instead to restrict the entry of any potential competitors. However, producers also want as large a market of consumers for their products as possible. Consumers of these products (buyers) should prefer free and open competition with as many producers as possible. However, consumers also want as few other consumers as possible for the products they buy. Note well that the interests of producers and consumers are diametrically opposed. This simple truth means that it will almost assuredly be impossible for any change in economic conditions, arising either out of natural dynamic forces in the economy or as a result of government policies, to be in the best interests of everyone in the country.

KEY TAKEAWAYS

- Greater competition (more sellers) in a market reduces the price of that good and lowers the well-being of the previous sellers. (Sellers dislike more sellers of the goods they sell.)
- Greater competition (more sellers) in a market raises the price of the buyer's goods and increases the well-being of the previous buyers. (Buyers like more sellers of the goods they buy.)
- The changes described above assume individuals are profit seeking.

EXERCISE

1. Consider two farmers, one with an endowment of five pounds of peaches, the other with an endowment of five pounds of cherries. Suppose these two farmers meet daily and make a mutually agreeable exchange of two pounds of peaches for three pounds of cherries.
 - a. Write down an expression for the terms of trade. Explain how the terms of trade relates to the dollar prices of the two goods.

Consider the following shocks (or changes). Explain how each of these shocks may influence the terms of trade between the farmers. Assume that each farmer's sole interest is to maximize her own utility.

- b. The cherry farmer arrives at the market with five extra pounds of cherries.
- c. The peach farmer has just finished reading a book titled *How to Influence People*.
- d. Damp weather causes mold to grow on 40 percent of the peaches.
- e. News reports indicate that cherry consumption can reduce the risk of cancer.

3.5 Three Traders with International Trade

LEARNING OBJECTIVE

1. Learn how international trade with competitor firms affects the distribution of income.

The farmer story can be placed in an international trade context with a simple adjustment. If we assume that Farmer Kim is from Korea, then the exchanges that take place in the second week reflect trade between countries. Farmer Smith's trade of oranges for apples with Farmer Kim represents U.S. exports of oranges in exchange for imports of apples from Korea. In the previous week, Farmer Kim was not present, thus all trade took place domestically. The change from week one to week two corresponds to a country moving from autarky to free trade.

Now consider the effects of trade in the United States. International trade makes Farmer Smith better off and Farmer Jones worse off compared to autarky. The critical point here is that free trade does not improve the well-being of everyone in the economy. Some individuals lose from trade.

We can characterize the winners and losers in a trade context by noting the relationship of the farmers to the trade pattern. Farmer Smith is an exporter of oranges. Farmer Jones must compete with imports on sales to Smith, thus we call Jones an import competitor. Our conclusion, then, is that export industries will benefit from free trade, while import-competing industries will suffer losses from free trade.

This result corresponds nicely with observations in the world. Generally, the most outspoken advocates of protection are the import-competing industries, while the avid free trade supporters tend to be affiliated with the export industries. In the United States, it is usually the importing textile, steel, and automobile industries calling for protection, while exporting companies like Boeing and Microsoft and the film industry preach the virtues of free trade.

KEY TAKEAWAYS

- Because export industries find more buyers for their products with international trade, export industries benefit from trade.
- Because trade increases the number of competitors import-competing industries face, trade harms import-competing industries.

EXERCISE

1. Choose a country. On the Internet, find the main exports and imports for that country and use this to indicate which industries are the likely winners and losers from trade.

3.6 The Nondiscrimination Argument for Free Trade

LEARNING OBJECTIVE

1. Learn how the constraint that trade policies be nondiscriminatory can lead people to choose free trade.

Each person has two roles in an economy: he or she is the maker and seller of some goods or services and the buyer of other goods and services. Most people work in a single industry. That means that each person's seller interest is rather limited. A steelworker's industry sells steel. A garment worker's industry sells clothes. A realtor sells realty services. Although some people may hold several jobs in different industries, most of the time a worker's income is tied to one particular industry and the products that industry sells. At the same time, most people's buying interests are quite diverse. Most individuals purchase hundreds of products every week—from food, books, and movies to cellular service, housing, and insurance.

We learned that it is in the best interests of sellers of goods to have as few other sellers of similar products as possible. We also learned that it is in the interests of buyers to have as many sellers of the goods they buy as possible. We can use this information to identify the very best economic situation for an individual with both buyer and seller interests.

Consider a worker in the insurance industry. This worker's income would be higher the less competition there was in the insurance sector. In the best of all circumstances, this worker's income would be the highest if his firm were a monopoly. However, as a buyer or consumer, this person would purchase hundreds or thousands of different products over the year. One such product would be clothing. The best situation here would be for all these products to be sold in markets with extensive competition—we might say perfect competition—since this would reduce the prices of the products he buys. Thus a monopoly in your own industry but perfect competition everywhere else is best from the individual's perspective.

However, consider a worker in the clothing industry. She too would be best served with a monopoly in her own industry and perfect competition everywhere else. But for her, the monopoly would have to be in the clothing sector, while everything else would need to be competitive.

Every country has workers in many different industries. Each one of these workers would be best served with a monopoly in his or her own industry and competition everywhere else. But clearly this is impossible unless the country produces only one good and imports everything else—something that's highly unlikely. That means there is no way for a government to satisfy everyone's interests by regulating competition.

However, we could demand that the government implement competition policies to satisfy one simple rule: nondiscrimination. Suppose we demand that the government treat everyone equally. Nondiscrimination rules out the scenarios benefiting individual workers. To allow steel to have a monopoly but to force competition in the clothing industry favors the steelworker at the expense of the clothing worker. The same applies if you allow a monopoly in the clothing industry but force competition in the steel sector.

Nondiscrimination would allow for only two competition policies in the extreme: either regulate so that all industries have a monopoly or regulate so that all industries face perfect competition. In terms of international trade policy, the nondiscriminatory options are either to allow free trade and open competition or to restrict trade equally by imposing tariffs that are so high that they completely restrict imports in every industry.

If people were forced to choose from the set of nondiscriminatory policies only, what would they choose? For every worker, there are plusses and minuses to each outcome. For the steelworker, for example, heavy protectionism would reduce competition in steel and raise his income. However, protectionism would also raise the prices of all the products he buys since competition would be reduced in all those industries as well. In short, protectionism means high income and high prices.

In contrast, free trade would mean the steel industry would face competition and thus steelworkers would get lower wages. However, all the goods the steelworker buys would be sold in more competitive markets and would therefore have lower prices. In short, the free trade scenario means low income and low prices.

6. The extent to which economic resources are transformed into products generating utility to consumers. Efficiency improves whenever greater output occurs per unit of input or when more satisfying consumption bundles are obtained.

So which nondiscriminatory outcome is better for a typical worker: high income and high prices or low income and low prices? Well, the Ricardian model in [Chapter 2 "The Ricardian Theory of Comparative Advantage"](#) and other models of trade provide an answer. Those models show that when free trade prevails, countries will tend to specialize in their comparative advantage goods, which will cause an overall increase in production. In other words, free trade promotes **economic efficiency**⁶. There will be more goods and services to be distributed to people under free trade

than there would be with no trade. Since the no-trade scenario corresponds to the protectionist choice, this outcome would leave people with fewer goods and services overall.

This means that the high-income and high-price scenario would leave people worse off than the low-income and low-price scenario. If people were well informed about these two outcomes and if they were asked to choose between these two nondiscriminatory policies, it seems reasonable to expect people would choose free trade. It is not hard to explain why a lower income might be tolerable as long as the prices of the hundreds of goods and services you purchase are low. Also, despite having the higher income with protection, what good is that if the prices of all the goods and services you purchase are also much higher?

Of course, there are also some intermediate nondiscriminatory trade policies the government could choose. For example, the government could do what Chile does and set a uniform tariff; Chile's is 6 percent currently. This would offer the same level of protection, or the same degree of restriction of competition, to all import-competing industries. However, since this would just be intermediate between the overall net benefits of free trade and the benefits of complete protection, the effects will be intermediate as well. Even with these options, then, the best nondiscriminatory choice to make is free trade.

KEY TAKEAWAYS

- Nondiscriminatory trade policies involve setting the same tariff on all imported products. The two extreme cases are either zero tariffs (free trade), or prohibitive tariffs (no trade).
- A free trade policy will cause lower income for each worker but also lower prices for all the goods and services purchased.
- A protectionist policy will cause higher incomes but also high prices for all the goods and services purchased.
- Given the choice between high income and high prices or low income and low prices, monopoly concerns suggest the latter would be chosen.

EXERCISE

1. Look at an individual country's bound tariff rates at the World Trade Organization (WTO). These can be found on the country pages of the WTO Web site. Go to http://www.wto.org/english/the WTO_e/whatis_e/tif_e/org6_e.htm, click on any country on the page, scroll down to the "Bound Tariffs" link, and click. It will load a PDF file with all the country's maximum tariffs.

Choose a country and determine whether the country applies discriminatory trade policies. If it does, identify several products that are highly protected and several that are not protected.

Chapter 4

Factor Mobility and Income Redistribution

This chapter continues the theme of income redistribution as a consequence of international trade. The focus here is the effect of factor immobility. In the Ricardian model presented in Chapter 2, it is assumed that workers can move freely and costlessly to another industry. In addition, it is assumed that each worker has the same productivity as every other worker in every other industry. This assumption makes it inconsequential if one industry shuts down because, if it does, the workers simply move to another industry where they will be just as productive and will likely earn a higher wage.

This chapter asks, “What happens if free and costless factor mobility does not hold?” The answer is provided by the results of the immobile factor model. This model is helpful for two important reasons. First, from a practical perspective, the model provides a reason why there can be both winners and losers as a result of international trade. Second, the model highlights an important technique used in economic analysis. Because the immobile factor model is identical to the Ricardian model in all but one assumption, the model demonstrates how changes in model assumptions directly impact the model implications and results. This is an important lesson about the method of economic analysis more generally.

4.1 Factor Mobility Overview

LEARNING OBJECTIVE

1. Identify the three dimensions across which factors of production may be mobile.

Factor mobility¹ refers to the ability to move factors of production—labor, capital, or land—out of one production process into another. Factor mobility may involve the movement of factors between firms within an industry, as when one steel plant closes but sells its production equipment to another steel firm. Mobility may involve the movement of factors across industries within a country, as when a worker leaves employment at a textile firm and begins work at an automobile factory. Finally, mobility may involve the movement of factors between countries either within industries or across industries, as when a farm worker migrates to another country or when a factory is moved abroad.

The standard assumptions in the trade literature are that factors of production are freely (i.e., without obstruction) and costlessly mobile between firms within an industry and between industries within a country but are immobile between countries.

The rationale for the first assumption—that factors are freely mobile within an industry—is perhaps closest to reality. The skills acquired by workers and the productivity of capital are likely to be very similar across firms producing identical or closely substitutable products. Although there would likely be some transition costs incurred, such as search, transportation, and transaction costs, it remains reasonable to assume for simplicity that the transfer is costless. As a result, this assumption is rarely relaxed.

The assumption that factors are easily movable across industries within a country is somewhat unrealistic, especially in the short run. Indeed, this assumption has been a standard source of criticism for traditional trade models. In the Ricardian and Heckscher-Ohlin models, factors are assumed to be homogeneous and freely and costlessly mobile between industries. When changes occur in the economy requiring the expansion of one industry and the contraction of another, it just happens. There are no search, transportation, or transaction costs. There is no unemployment of resources. Also, since the factors are assumed to be homogeneous, once transferred to a completely different industry, they

1. The ability to move factors of production—labor, capital, or land—out of one production process and into another.

immediately become just as productive as the factors that had originally been employed in that industry. Clearly, these conditions cannot be expected to hold in very many realistic situations. For some, this inconsistency is enough to cast doubt on all the propositions that result from these theories.

It is important to note, however, that trade theory has attempted to deal with this concern to some extent. The immobile factor model (in [Chapter 4 "Factor Mobility and Income Redistribution"](#)) and the specific factor model (in [Chapter 5 "The Heckscher-Ohlin \(Factor Proportions\) Model", Section 5.15 "The Specific Factor Model: Overview"](#)) represent attempts to incorporate factor immobility precisely because of the concerns just mentioned. Although these models do not introduce resource transition in a complicated way, they do demonstrate important income redistribution results and allow one to infer the likely effects of more complex adjustment processes by piecing together the results of several models. (See [Chapter 5 "The Heckscher-Ohlin \(Factor Proportions\) Model", Section 5.17 "Dynamic Income Redistribution and Trade"](#), especially.)

Another important aspect of factor mobility involves the mobility of factors between countries. In most international trade models, factors are assumed to be immobile across borders. Traditionally, most workers remain in their country of national origin due to immigration restrictions, while government controls on capital have in some periods restricted international movements of capital. When international factor mobility is not possible, trade models demonstrate how national gains can arise through trade in goods and services.

Of course, international mobility can and does happen to varying degrees. Workers migrate across borders, sometimes in violation of immigration laws, while capital flows readily across borders in today's markets. The implications of international factor mobility have been addressed in the context of some trade models. A classic result by Robert A. Mundell (1957) demonstrates that international factor mobility can act as a substitute for international trade in goods and services. In other words, to realize all the gains from international exchange and globalization, countries need to *either* trade freely *or* allow factors to move freely between countries. Robert A. Mundell, "International Trade and Factor Mobility," *American Economic Review* 47 (1957): 321–35. It is not necessary to have both. Mundell's result contradicts a popular argument that free trade can only benefit countries if they also allow workers to move freely across borders.

KEY TAKEAWAYS

- Factors of production are potentially mobile in three distinct ways:
 - Between firms within the same industry
 - Between industries within the same country
 - Between firms or industries across countries
- A standard simplifying assumption in many trade models is that factors of production are freely and costlessly mobile between firms and between industries but not between countries.
- The immobile factor model and the specific factor model are two models that assume a degree of factor *immobility* between industries.

EXERCISES

1. Name several impediments to the free movement of workers between two industries.
2. Name several costs associated with the movement of workers between two industries.

4.2 Domestic Factor Mobility

LEARNING OBJECTIVE

- Understand how the different types of factors display different degrees of factor mobility.

Domestic factor mobility² refers to the ease with which productive factors like labor, capital, land, natural resources, and so on can be reallocated across sectors within the domestic economy. Different degrees of mobility arise because there are different costs associated with moving factors between industries.

As an example of how the adjustment costs vary across factors as factors move between industries, consider a hypothetical textile firm that is going out of business.

The textile firm employs a variety of workers with different types of specialized skills. One of these workers is an accountant. Fortunately for the accountant, she has skills that are used by all businesses. Although there may be certain specific accounting techniques associated with the textile industry, it is likely that this worker could find employment in a variety of industries. The worker would still suffer some adjustment costs such as a short-term reduction in salary, search costs to find another job, and the anxiety associated with job loss. However, assuming there is no glut of accountants in the economy, this worker is likely to be fairly mobile.

Consider another worker who is employed as a seamstress in the textile firm. If the textile industry as a whole is downsizing, then it is unlikely that she will find a job in another textile plant. Also, the skills of a seamstress are not widely used in other industries. For this worker, finding another job may be very difficult. It may require costs beyond those incurred by the accountant. This worker may decide to learn a new profession by attending a vocational school or going to college. All of this requires more time and incurs a greater cost.

- When productive factors like labor, capital, land, natural resources, and so on can be reallocated across sectors within a domestic economy.

Next consider the capital equipment used in the textile plant. The looms that are used to weave cloth are unlikely to be very useful or productive in any other industry. Remaining textile firms might purchase them, but only if the prices are

very low. Ultimately, these machines are likely to fall into disuse and be discarded. Looms exhibit very low mobility to other industries.

However, consider a light truck owned and operated by the firm. This truck could easily be sold and used by another firm in a completely different industry. The only costs would be the cost of making the sale (advertisements, sales contracts, etc.) and perhaps the cost of relabeling the truck with the new company name. The truck is relatively costlessly transferable across industries.

Finally, consider the land on which the textile plant operates. Depending on the location of the firm and the degree of new business creations or expansions in the area, the land may or may not be transferred easily. One possible outcome is that the property could be sold to another business that would recondition it to suit its needs. In this case, the cost of mobility includes the transactions costs to complete the sale plus the renovation costs to fix up the property for its new use. Alternatively, the land could remain for sale for a very long time during which the plant merely becomes an eyesore. In this case, the land's immobility may last for years.

These examples suggest that the cost of factor mobility varies widely across factors of production. Some factors such as accountants and trucks may be relatively costless to move. Other factors like looms and seamstresses may be very costly to move. Some factors like land may be easy to move in some instances but not in others.

KEY TAKEAWAY

- The ability and cost of factor mobility across industries depends largely on how widespread the demands are for that particular factor.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Between *truck driver* and *bricklayer*, this occupation is likely to be more easily adapted for use in an alternative industry.
 - b. Between *accountant* and *robotics engineer*, this occupation is likely to be more easily adapted for use in an alternative industry.
 - c. Between *professional baseball player* and *chemist*, this occupation is likely to be more easily adapted for use in an alternative industry.
2. Suppose a chemist loses her job at a pharmaceutical company. What other industries are most likely to demand the services of a chemist? What other industries are least likely to demand the services of a chemist?

4.3 Time and Factor Mobility

LEARNING OBJECTIVE

1. Learn why time passage is a very important element affecting a particular factor's mobility across industries.

The degree of mobility of factors across industries is greatly affected by the passage of time. In the very, very short run—say, over a few weeks' time—most unemployed factors are difficult to move to another industry. Even the worker whose skills are readily adaptable to a variety of industries would still have to take time to search for a new job. Alternatively, a worker in high demand in another industry might arrange for a brief vacation between jobs. This means that over the very short run, almost all factors are relatively immobile.

As time passes, the most mobile factors begin to find employment in other industries. At the closed textile plant, some of the managers, the accountants, and some others may find new jobs within four to six months. The usable capital equipment may be sold to other firms. Looms in good working condition may be bought by other textile plants still operating. Trucks and other transport equipment will be bought by firms in other industries. As time progresses, more and more factors find employment elsewhere.

But what about the seamstress near retirement whose skills are not in demand and who is unwilling to incur the cost of retraining? Or the capital equipment that is too old, too outdated, or just inapplicable elsewhere in the economy? These factors, too, can be moved to other industries given enough time. The older workers will eventually retire from the workforce. Their replacements will be their grandchildren, who are unlikely to seek the skills or jobs of their grandparents.

Merely recall the decline of family farms in America. For generations, children followed parents as farmers until it eventually became unprofitable to continue to operate the same way. As the number of farmers declined, the children of farmers began to move into the towns and cities. They went to colleges and often learned skills very different from their parents and grandparents.

In this way, as generations age and retire, the children acquire the new skills in demand in the modern economy, and the distribution of skills in the workforce

changes. Labor automatically becomes mobile across industries if we allow enough time to pass.

Consider also the capital equipment that is unusable in any other industry. This capital is also mobile in a strange sort of way. Generally, as capital equipment is used, its value declines. Often the cost of repairs rises for an older machine. Older machines may be less productive than newer models, also reducing their relative worth. When capital depreciates, or loses its value, sufficiently, a firm continuing to produce would likely invest in a new machine. Investment requires the owners of the firm to forgo profits in order to purchase new capital equipment.

Now suppose the firm is a textile plant and the owners are shutting it down. The capital equipment at the firm will suddenly depreciate more rapidly than originally anticipated.

As this equipment depreciates, however, new investments will not be directed at the same type of capital. Instead, investors will purchase different types of capital that have the potential for profits in other industries. In this way, over time, as the current capital stock depreciates, new investment is made in the types of capital needed for production in the future. With enough time, the capital stock is moved out of declining, unprofitable industries and into expanding, profitable industries.

In summary, virtually all factors are immobile across industries in the very short run. As time progresses and at some cost of adjustment, factors become mobile across sectors of the economy. Some factors move more readily and at less cost than others. In the long run, all factors are mobile at some cost. For workers, complete mobility may require the passing of a generation out of the workforce. For capital, complete mobility requires depreciation of the unproductive capital stock, followed by new investment in profitable capital.

KEY TAKEAWAY

- The ability of a factor to find employment in a new industry tends to increase as time passes.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Between *short run* and *long run*, this time frame is more associated with unlimited factor mobility.
 - b. The term used to describe the fact that machines wear out over time.
 - c. Of 10 percent, 50 percent, or 100 percent, this is the more likely percentage of production factors that can adjust between diverse industries in the short run.
 - d. Of 10 percent, 50 percent, or 100 percent, this is the more likely percentage of production factors that can adjust between diverse industries in the long run.
 - e. The term used to describe the period of time in which production factors cannot move between industries within a country.

4.4 Immobile Factor Model Overview and Assumptions

LEARNING OBJECTIVES

1. Learn how the immobile factor model differs from the Ricardian model.
2. Learn the assumptions of a standard immobile factor trade model.

Overview

The **immobile factor model**³ highlights the effects of factor immobility between industries within a country when a country moves to free trade. The model is the standard Ricardian model with one variation in its assumptions. Whereas in the Ricardian model, labor can move costlessly between industries, in the immobile factor model, we assume that the cost of moving a factor is prohibitive. This implies that labor, the only factor, remains stuck in its original industry as the country moves from autarky to free trade.

The assumption of labor immobility allows us to assess the short-run impact of movements to free trade where the short run is defined as the period of time when all factors of production are incapable of moving between sectors. The main result of the model is that free trade will cause a redistribution of income such that some workers gain from trade, while others lose from trade.

Assumptions

The immobile factor model assumptions are identical to the Ricardian model assumptions with one exception. In this model, we assume that L_C and L_W are exogenous. This means that there is a fixed supply of cheese workers and wine workers. Cheese workers know how to make cheese but cannot be used productively in the wine industry, and wine workers cannot be used productively in the cheese industry. This assumption differs from the Ricardian model, which assumed that labor was freely mobile across industries. In the Ricardian model, a cheese worker who moved to the wine industry would be immediately as productive as a longtime wine worker.

- 3. A standard Ricardian model with one variation in its assumptions, namely, that labor, the sole factor of production, is immobile between industries within a country.
- 4. Factors that can be moved by their owners to another production process without impediment and without incurring any adjustment costs.

Neither assumption—**free and costless mobility**⁴ nor complete immobility—is entirely realistic. Instead, they represent two extreme situations. The Ricardian assumption can be interpreted as a long-run scenario. Given enough time, all factors can be moved and become productive in other industries. The immobile

factor assumption represents an extreme short-run scenario. In the very short run, it is difficult for any factor to be moved and become productive in another industry. By understanding the effects of these two extremes, we can better understand what effects to expect in the real world, characterized by incomplete and variable factor mobility.

What follows is a description of the standard assumptions in the immobile factor model. We assume perfect competition prevails in all markets.

Number of Countries

The model assumes two countries to simplify the model analysis. Let one country be the United States, the other France. Note that anything related exclusively to France in the model will be marked with an asterisk.

Number of Goods

The model assumes there are two goods produced by both countries. We assume a barter economy. This means that no money is used to make transactions. Instead, for trade to occur, goods must be traded for other goods. Thus we need at least two goods in the model. Let the two produced goods be wine and cheese.

Number of Factors

The model assumes there are two factors of production used to produce wine and cheese. Wine production requires wine workers, while cheese production requires cheese workers. Although each of these factors is a kind of labor, they are different types because their productivities differ across industries.

Consumer Behavior

Factor owners are also the consumers of the goods. We assume the factor owners have a well-defined utility function defined over the two goods. Consumers maximize utility to allocate income between the two goods.

A General Equilibrium

The immobile factor model is a general equilibrium model. The income earned by the factor is used to purchase the two goods. The industries' revenue in turn is used to pay for the factor services. The prices of the outputs and the factor are determined such that supply and demand are equalized in all markets simultaneously.

Demand

We will assume that aggregate demand is homothetic in this model. This implies that the marginal rate of substitution between the two goods is constant along a ray from the origin. We will assume further that aggregate demand is identical in both of the trading countries. Note that this assumption is a technical detail that affects how the trading equilibrium is depicted but is not very important in understanding the main results.

Supply

The production functions in [Table 4.1 "Production of Cheese"](#) and [Table 4.2 "Production of Wine"](#) represent industry production, not firm production. The industry consists of many small firms in light of the assumption of perfect competition.

Table 4.1 Production of Cheese

United States	France
$Q_C = \frac{\bar{L}_C [\text{hrs}]}{a_{LC} \left[\frac{\text{hrs}}{\text{lb}} \right]}$	$Q_C^* = \frac{\bar{L}_C^*}{a_{LC}^*}$

where

Q_C = quantity of cheese produced in the United States

\bar{L}_C = fixed amount of labor applied to cheese production in the United States

a_{LC} = unit labor requirement in cheese production in the United States (hours of labor necessary to produce one unit of cheese)

*All starred variables are defined in the same way but refer to the production process in France.

Table 4.2 Production of Wine

United States	France
$Q_W = \frac{\bar{L}_W [\text{hrs}]}{a_{LW} \left[\frac{\text{hrs}}{\text{gal}} \right]}$	$Q_W^* = \frac{\bar{L}_W^*}{a_{LW}^*}$

where

Q_W = quantity of wine produced in the United States

\bar{L}_W = amount of labor applied to wine production in the United States

a_{LW} = unit labor requirement in wine production in the United States (hours of labor necessary to produce one unit of wine)

*All starred variables are defined in the same way but refer to the production process in France.

The unit labor requirements define the technology of production in the two countries. Differences in these labor costs across countries represent differences in technology.

KEY TAKEAWAY

- The immobile factor model is a two-country, two-good, two-factor, perfectly competitive general equilibrium model that is identical to the Ricardian model except that labor cannot move across industries.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The assumption that distinguishes the immobile factor model from the Ricardian model.
 - b. The term describing the period of time encompassed by the immobile factor model.
 - c. The firms’ objective in the immobile factor model.
 - d. The consumers’ objective in the immobile factor model.
 - e. The term for the entire collection of assumptions made in the immobile factor model.

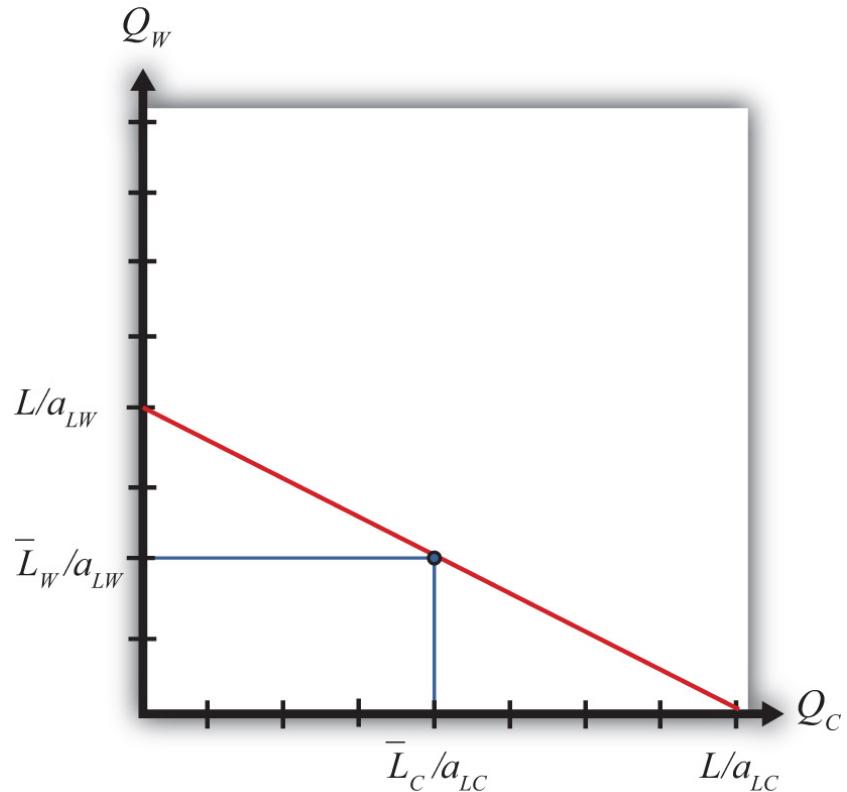
4.5 The Production Possibility Frontier in the Immobile Factor Model

LEARNING OBJECTIVE

1. Learn how the immobile factor model's production possibility frontier (PPF) is drawn and how it compares with the Ricardian model's PPF.

To derive the production possibility frontier (PPF) in the immobile factor model, it is useful to begin with a PPF from the Ricardian model. In the Ricardian model, the PPF is drawn as a straight line with endpoints given by L/a_{LC} and L/a_{LW} , where L is the total labor endowment available for use in the two industries (see [Figure 4.1 "The Immobile Factor Model PPF"](#)). Since labor is moveable across industries, any point along the PPF is a feasible production point that maintains full employment of labor.

Figure 4.1 The Immobile Factor Model PPF



Next, let's suppose that some fraction of the L workers are cheesemakers, while the remainder are winemakers. Let \bar{L}_C be the number of cheesemakers and \bar{L}_W be the number of winemakers such that $\bar{L}_C + \bar{L}_W = L$. If we assume that these workers cannot be moved to the other industry, then we are in the context of the immobile factor model.

In the immobile factor model, the PPF reduces to a single point represented by the blue dot in [Figure 4.1 "The Immobile Factor Model PPF"](#). This is the only production point that generates full employment of both wine workers and cheese workers. The production possibility set (PPS) consists of the set of points that is feasible whether or not full employment is maintained. The PPS is represented by the rectangle formed by the blue lines and the Q_C and Q_W axes.

Notice that in the immobile factor model, the concept of opportunity cost is not defined because it is impossible, by assumption, to increase the output of either good. No opportunity cost also means that neither country has a comparative advantage as defined in the Ricardian model. However, this does not mean there is no potential for advantageous trade.

KEY TAKEAWAYS

- The PPF in an immobile factor model consists of a single point because a fixed labor supply in each industry leads to a fixed quantity of each good that can be produced with full employment.
- Opportunity cost is not defined in the immobile factor model.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. A description of the production possibility set in the immobile factor model.
 - b. Of *true or false*, the opportunity cost of cheese production is not defined in the immobile factor model.
 - c. Of *true or false*, the production point $(0, 0)$ is a part of the production possibility set in the immobile factor model.
 - d. Of *true or false*, the production point $(0, 0)$ is a part of the production possibility frontier in the immobile factor model.

4.6 Autarky Equilibrium in the Immobile Factor Model

LEARNING OBJECTIVES

1. Depict an autarky equilibrium in the immobile factor model.
2. Determine the autarky terms of trade given particular assumptions concerning technology, endowments, and demands.

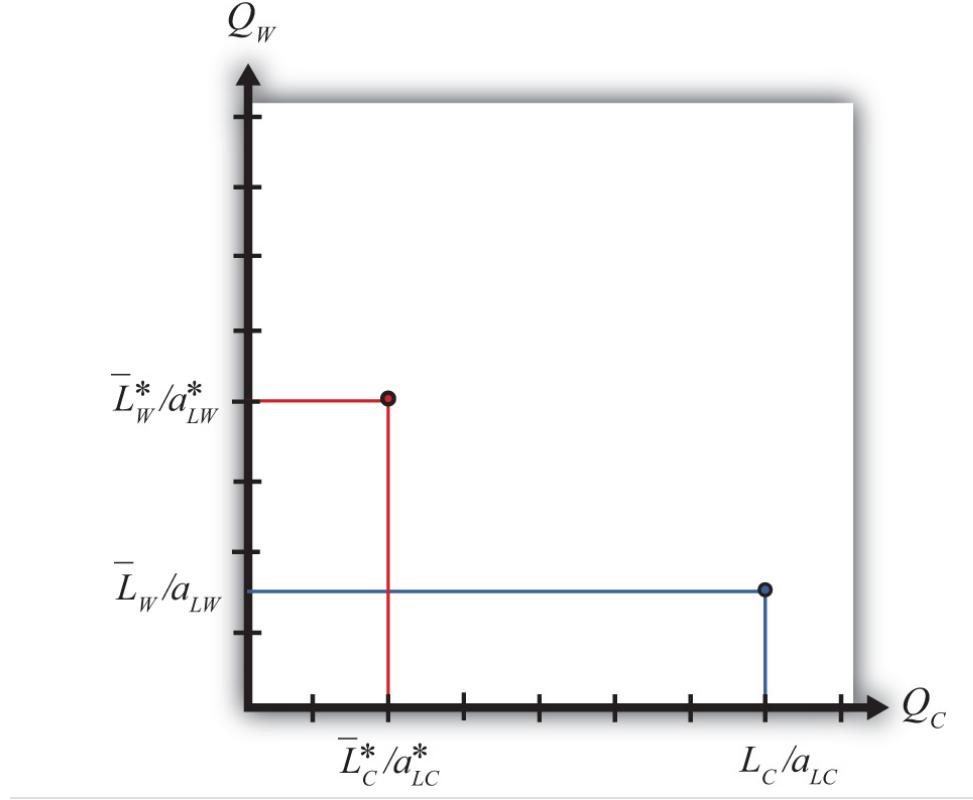
Suppose two countries, the United States and France, have the exactly the same number of winemakers and cheesemakers. This means $\bar{L}_C = \bar{L}_C^*$ and $\bar{L}_W = \bar{L}_W^*$. Suppose also that the United States has an absolute advantage in the production of cheese, while France has the absolute advantage in the production of wine. This means $a_{LC} < a_{LC}^*$ and $a_{LW}^* < a_{LW}$. Also, assume that the preferences for the two goods in both countries are identical.

For simplicity, let aggregate preferences be represented by a homothetic utility function. These functions have the property that for any price ratio, the ratio of the two goods consumed is equal to a constant. One function with this property is $\frac{Q_W^D}{Q_C^D} = \frac{P_C}{P_W}$, where Q_C^D is the aggregate quantity of cheese demanded and Q_W^D is the aggregate quantity of wine demanded. This function says that the ratio of the quantity of wine demanded to the quantity of cheese demanded must equal the price ratio.

For example, suppose that consumers face a price ratio $P_C/P_W = 2$ gallons of wine per pound of cheese. In this case, consumers will demand wine to cheese in the same ratio: two gallons per pound. Suppose the price ratio rises to $P_C/P_W = 3$. This means that cheese becomes more expensive than wine. At the higher price ratio, consumers will now demand three gallons of wine per pound of cheese. Thus as the relative price of cheese rises, the relative demand for wine rises as consumers substitute less expensive wine for more expensive cheese. Similarly, as the price of wine falls, the relative demand for wine rises.

The PPFs for the two countries in this case are plotted in [Figure 4.2 "The U.S. and France's PPFs"](#). The United States produces more cheese than France, while France produces more wine than the United States. Because the factors are immobile, the ratio of wine to cheese production in the United States must be $\frac{Q_W}{Q_C} = \frac{\bar{L}_W/a_{LW}}{\bar{L}_C/a_{LC}}$.

Figure 4.2 The U.S. and France's PPFs



In autarky, the quantity demanded of each good must equal the quantity supplied. This implies that the ratios of quantities must also be equalized such that

$$\frac{Q_W^D}{Q_C^D} = \frac{Q_W}{Q_C}.$$

Substituting from above yields the autarky price ratio in the United States:

$$\left(\frac{P_C}{P_W} \right)_{Aut} = \frac{\bar{L}_W/a_{LW}}{\bar{L}_C/a_{LC}} = \frac{a_{LC}}{a_{LW}} \frac{\bar{L}_W}{\bar{L}_C}.$$

Similarly, France's autarky price ratio is the following:

$$\left(\frac{P_C^*}{P_W^*} \right)_{Aut} = \frac{a_{LC}^*}{a_{LW}^*} \frac{\bar{L}_W^*}{\bar{L}_C^*}.$$

Since by assumption the two countries have identical labor endowments, the United States has an absolute advantage in cheese production, and France has an absolute advantage in wine production, it follows that

$$\left(\frac{P_C}{P_W} \right)_{Aut} < \left(\frac{P_C^*}{P_W^*} \right)_{Aut}.$$

Note that the same terms of trade relationship would follow if instead we assumed that the unit labor requirements, and hence the technologies, were the same in both countries but allowed the endowment of cheesemakers to be greater in the United States while the endowment of winemakers was larger in France.

In autarky, each country will produce at its production possibility point and, since there is no trade, will consume the same quantities of cheese and wine. The price of cheese is lower in the United States in autarky because it produces relatively more cheese than France given its absolute advantage, and that extra supply tends to force the price of cheese down relative to France. Similarly, France's absolute advantage in wine causes it to produce more wine than the United States, which causes the price of wine in France to be lower than in the United States.

KEY TAKEAWAYS

- In autarky, in the immobile factor model, consumption will occur at the only production point possible in the model.
- The autarky terms of trade for a good will be lower in the country with the productivity advantage (or the greater factor endowment in that product).

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This happens to the demand for cheese if the price ratio P_C/P_W rises.
 - b. This happens to the demand for cheese if one kilogram of cheese now trades for one liter of wine rather than two liters.
 - c. This happens to the demand for cheese if one liter of wine now trades for three kilograms of cheese rather than four kilograms.
 - d. With homothetic preferences, the ratio of consumer demands of wine to cheese will equal *this other ratio*.

4.7 Depicting a Free Trade Equilibrium in the Immobile Factor Model

LEARNING OBJECTIVE

1. Depict the production, consumption, and trade patterns for two countries in an immobile factor model in free trade.

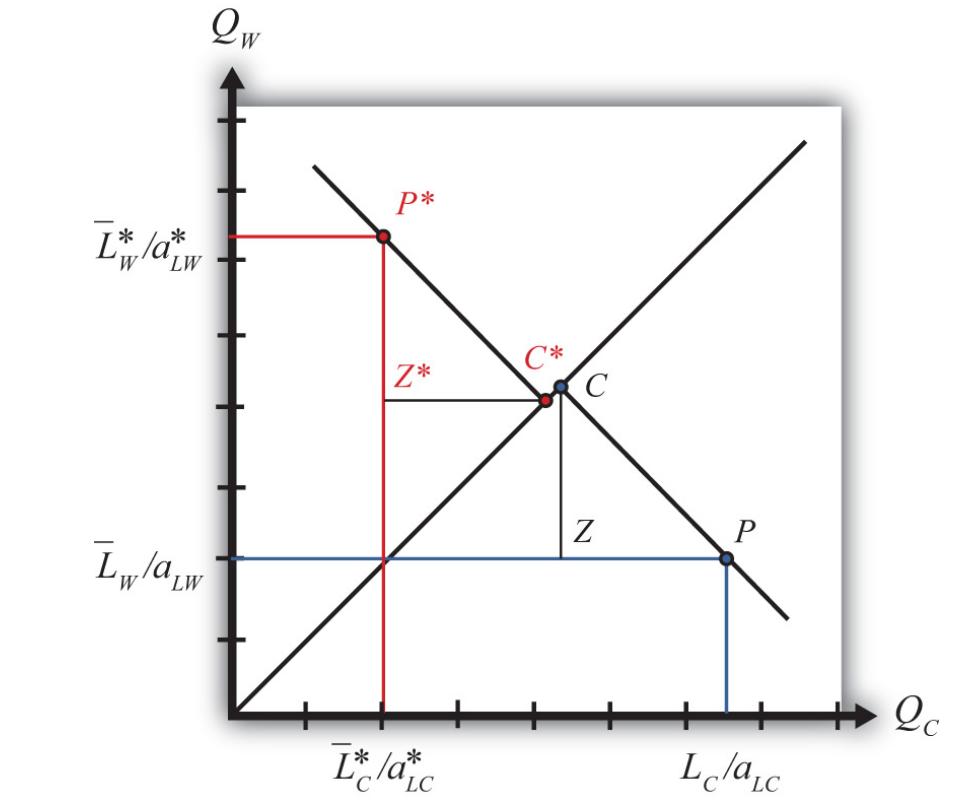
Differences in price ratios are all that's needed to stimulate trade once the barriers to trade are removed. Since the price of cheese is higher in France upon the opening of free trade, U.S. cheese producers will begin to export cheese to the French market, where they will make a greater profit. Similarly, French wine producers will export wine to the U.S. market, where it commands a higher price. The effect of the shift in supply is to force the price of cheese relative to wine down in France and up in the United States until they meet at a price ratio that equalizes world supply of wine and cheese with world demand for wine and cheese.

When a free trade equilibrium is reached, the following conditions will prevail:

1. Both countries face the same terms of trade: $(P_C/P_W)_{FT}$.
2. Both countries will demand the same ratio of wine to cheese: Q_W^D/Q_C^D .
3. Exports of cheese by the United States will equal imports of cheese by France.
4. Exports of wine by France will equal imports of wine by the United States.

The free trade equilibrium is depicted in [Figure 4.3 "A Free Trade Equilibrium in the Immobile Factor Model"](#). The countries produce at the points P^* and P and consume after trade at the points C^* and C , respectively. Thus the United States exports ZP units of cheese, while France imports the equivalent, C^*Z^* . Similarly, France exports Z^*P^* units of wine, while the United States imports the equivalent, CZ . Each country trades with the other in the ratio CZ/ZP gallons of wine per pound of cheese. This corresponds to the free trade price ratio, $(P_C/P_W)_{FT}$, represented by the slope of the lines C^*P^* and CP .

Figure 4.3 A Free Trade Equilibrium in the Immobile Factor Model



The equilibrium demonstrates that with trade both countries are able to consume at a point that lies outside their production possibility set (PPS). In other words, trade opens up options that were not available to the countries before.

KEY TAKEAWAY

- In an immobile factor model, free trade enables both countries to consume a mix of goods that were not available to them before trade.

EXERCISE

1. Suppose two countries, Brazil and Argentina, can be described by an immobile factor model. Assume they each produce wheat and chicken using labor as the only input. Suppose the two countries move from autarky to free trade with each other. Assume the terms of trade change in each country as indicated below. In the remaining boxes, indicate the effect of free trade on the variables listed in the first column in both Brazil and Argentina. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 4.3 EFFECTS OF FREE TRADE

	In Brazil	In Argentina
P_c/P_w	+	-
Output of Wheat		
Output of Chicken		
Exports of Wheat		
Imports of Wheat		

4.8 Effect of Trade on Real Wages

LEARNING OBJECTIVES

1. Learn how to measure real wages in the immobile factor model.
2. Learn how real wages change when a country moves from autarky to free trade.

We calculate real wages to determine whether there are any income redistribution effects in moving to free trade. The real wage formulas in the immobile factor model are the same as in the Ricardian model since perfect competition prevails in both industries. However, the wage paid to cheese workers no longer must be the same as the wage of wine workers. Cheese workers' wages could be higher since wine workers cannot shift to the cheese industry to take advantage of the higher wage.

When the countries move from autarky to free trade, the price ratio in the United States, P_C/P_W , rises.

The result is a redistribution of income as shown in Table 4.4 "Changes in Real Wages (Autarky to Free Trade): ". Cheese workers face no change in their real wage in terms of cheese and experience an increase in their real wage in terms of wine.

Table 4.4 Changes in Real Wages (Autarky to Free Trade): P_C/P_W Rises

	In Terms of Cheese	In Terms of Wine
Real Wage of U.S. Cheese Workers	$\frac{w_C}{P_C} = \frac{1}{a_{LC}}$ (no change)	$\frac{w_C}{P_C} = \frac{1}{a_{LC}} \frac{P_C}{P_W}$ (rises)
Real Wage of U.S. Wine Workers	$\frac{w_W}{P_C} = \frac{1}{a_{LW}} \frac{P_W}{P_C}$ (falls)	$\frac{w_W}{P_W} = \frac{1}{a_{LW}}$ (no change)
where		
P_C = price of cheese		

	In Terms of Cheese	In Terms of Wine
P_W = price of wine w_C = wage paid to cheese workers w_W = wage paid to wine workers a_{LC} = unit labor requirement in cheese production in the United States (hours of labor necessary to produce one unit of cheese) a_{LW} = unit labor requirement in wine production in the United States (hours of labor necessary to produce one unit of wine)		

Thus cheese workers are most likely better off in free trade. Wine workers face no change in their real wage in terms of wine but suffer a decrease in their real wage in terms of cheese. This means wine workers are likely to be worse off as a result of free trade.

Since one group of workers realizes real income gains while another set suffers real income losses, *free trade causes a redistribution of income within the economy. Free trade results in winners and losers in the immobile factor model.*

In France, the price ratio, P_C/P_W , falls when moving to free trade. The result is a redistribution of income similar to the United States as shown in [Table 4.5 "Changes in Real Wages \(Autarky to Free Trade\): "](#). Cheese workers face no change in their real wage in terms of cheese and experience a decrease in their real wage in terms of wine.

Table 4.5 Changes in Real Wages (Autarky to Free Trade): P_C/P_W Falls

	In Terms of Cheese	In Terms of Wine
Real Wage of French Cheese Workers	$\frac{w_C}{P_C} = \frac{1}{a_{LC}}$ (no change)	$\frac{w_C}{P_C} = \frac{1}{a_{LC}} \frac{P_C}{P_W}$ (falls)
Real Wage of French Wine Workers	$\frac{w_W}{P_C} = \frac{1}{a_{LW}} \frac{P_W}{P_C}$ (rises)	$\frac{w_W}{P_W} = \frac{1}{a_{LW}}$ (no change)

Thus cheese workers are most likely worse off in free trade. Wine workers face no change in their real wage in terms of wine but realize an increase in their real wage in terms of cheese. This means wine workers are likely to be better off as a result of free trade.

Since one group of workers realizes real income gains while another set suffers real income losses, free trade causes a redistribution of income within the economy. *Free trade results in winners and losers in both the United States and France.* In both countries, the winners are those workers who work in the industry whose output price rises, while the losers work in the industry whose output price falls. But because the price changes are due to the movement to free trade, it is also true that the output price increases occur in the export industries in both countries, while the price declines occur in the import-competing industries. Thus it follows that *a movement to free trade will benefit those workers who work in the export industry and harm those workers who work in the import-competing industry.*

KEY TAKEAWAYS

- When countries move to free trade and labor is immobile, in the export industry the real wage with respect to the exported good remains constant, but the real wage with respect to the import good rises in both countries.
- When countries move to free trade and labor is immobile, in the import industry the real wage with respect to the imported good remains constant, but the real wage with respect to the export good falls in both countries.
- When countries move to free trade and labor is immobile, in general, workers in the export industry benefit, while workers in the import-competing industry lose.

EXERCISES

1. According to an immobile factor model, which groups are likely to benefit very shortly after trade liberalization occurs? Which groups are likely to lose very shortly after trade liberalization occurs?
2. Suppose two countries, Brazil and Argentina, can be described by an immobile factor model. Assume they each produce wheat and chicken using labor as the only input. Suppose the two countries move from autarky to free trade with each other. Assume the terms of trade change in each country as indicated below. In the remaining boxes, indicate the effect of free trade on the variables listed in the first column in both Brazil and Argentina. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 4.6 REAL WAGE EFFECTS

	In Brazil	In Argentina
P_c/P_w	+	-
Real Wage of Chicken Workers in Terms of Chicken		
Real Wage of Chicken Workers in Terms of Wheat		
Real Wage of Wheat Workers in Terms of Chicken		
Real Wage of Wheat Workers in Terms of Wheat		

4.9 Intuition of Real Wage Effects

LEARNING OBJECTIVE

- Understand intuitively why real wages change differently in the immobile factor model.

When the United States and France move from autarky to free trade, the U.S. price of cheese rises and the United States begins to export cheese. The French price of wine rises and France begins to export wine. In both of these industries, the higher prices generate higher revenue, and since profits must remain equal to zero because of competition in the industry, higher wages are paid to the workers. As long as the factors remain immobile, other workers do not enter the higher wage industry, so these higher wages can be maintained. Thus in both countries *real wages rise for workers in the export industries*.

The movement from autarky to free trade also causes the price of wine to fall in the United States while the United States imports wine and the price of cheese to fall in France while France imports cheese. Lower prices reduce the revenue to the industry, and to maintain zero profit, wages are reduced proportionally. Since workers are assumed to be immobile, workers cannot flee the low-wage industry and thus low wages are maintained. Thus in both countries *real wages fall for workers in the import-competing industries*.

But isn't it possible for the owners of the firms in the export industries to claim all the extra revenue for themselves? In other words, maybe when the price rises the owners of the export firms simply pay the CEO and the rest of management a few extra million dollars and do not give any of the extra revenue to the ordinary workers. Actually, this is unlikely under the assumptions of the model. First of all, the model has no owners or management. Instead, all workers are assumed to be the same, and no workers have any special ownership rights. But let's suppose that there is an owner. The owner can't claim a huge pay increase because the industry is assumed to be perfectly competitive. This means that there are hundreds or thousands of other export firms that have all realized a price increase. Although workers are assumed to be immobile across industries, they are not immobile *between firms* within an industry.

So let's suppose that all the firm's owners simply pocket the extra revenue. If one of these owners wants to make even more money, it is now possible. All she must do is

reduce her pay somewhat and offer her workers a higher wage. The higher wage will entice other workers in the industry to move to the generous firm. By increasing workers' wages, this owner can expand her own firm's output at the expense of other firms in the industry. Despite a lower wage for the owner, as long as the increased output is sufficiently large, the owner will make even more money for herself than she would have had she not raised worker wages. However, these extra profits will only be temporary since other owners would soon be forced to raise worker wages to maintain their own output and profit. It is this competition within the industry that will force wages for workers up and the compensation for owners down. In the end, economic profit will be forced to zero. Zero economic profit assures that owners will receive just enough to prevent them from moving to another industry.

KEY TAKEAWAYS

- The assumption of immobile labor means that workers cannot take advantage of higher wages paid in another industry after opening to trade. Lack of competition in the labor market allows export industry wages to rise and import-competing industry wages to fall.
- Competition between firms within an industry assures that all workers receive an identical wage and no one group within the industry can enjoy above-normal profit in the long run.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is "a tax on imports," then the correct question is "What is a tariff?"
 - a. Of *true or false*, factors can move freely and costlessly between industries in an immobile factor model.
 - b. Of *true or false*, factors can move freely and costlessly between firms within an industry in an immobile factor model.

4.10 Interpreting the Welfare Effects

LEARNING OBJECTIVE

1. Understand how national welfare is affected by free trade in an immobile factor model and why compensation cannot assure everyone gains.

The real wage calculations show that some workers gain from trade, while others lose from trade. On the other hand, we showed that the economy is able to jump to a higher aggregate indifference as a result of free trade. The increase in aggregate welfare is attributable entirely to an increase in consumption efficiency. A reasonable question to ask at this juncture is whether the winners from trade could compensate the losers such that every worker is left no worse off from free trade. The answer to this question is no in the context of this model.

In the immobile factor model, there is no increase in world productive efficiency. The immobility of factors implies that world output is the same with trade as it was in autarky. This means that the best that compensation could provide is to return everyone to their autarky consumption levels. And the only way to do that is to eliminate trade. There simply is no way to increase the total consumption of each good for every worker after trade begins.

Sometimes economists argue that since the model displays an increase in consumption efficiency, this means that the country is better off with trade. While technically this is true, it is important to realize that statements about what's best for a country in the aggregate typically mask the effects on particular individuals. The immobile factor model suggests that in the very short run, movements to free trade will very likely result in a redistribution of income with some groups of individuals suffering real income losses. It will be very difficult to convince those who will lose that free trade is a good idea because the aggregate effects are positive.

Furthermore, since there is no way for the winners to compensate the losers such that everyone gains, the model implies that the movement to free trade can be a zero-sum game, at least in the very short run. This means that the sum of the gains to the winners is exactly equal to the sum of the losses to the losers.

In the Heckscher-Ohlin model, we will show that income redistribution is possible even in the long run when an economy moves to free trade. However, in that case, free trade will be a positive-sum game in that the sum of the gains will exceed the sum of the losses.

KEY TAKEAWAY

- In the immobile factor model, because there is no increase in output of either good when moving to free trade, there is no way for compensation to make everyone better off after trade.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, this is what happens to the output of cheese in France in an immobile factor model when it moves to free trade.
 - b. Of *increase, decrease, or stay the same*, this is what happens to the output of wine in France in an immobile factor model when it moves to free trade.
 - c. Of *increase, decrease, or stay the same*, this is what happens to world productive efficiency in an immobile factor model when two countries move to free trade.
 - d. Of *true or false*, compensation provided to the losers from trade can assure that everyone gains from trade in an immobile factor model.

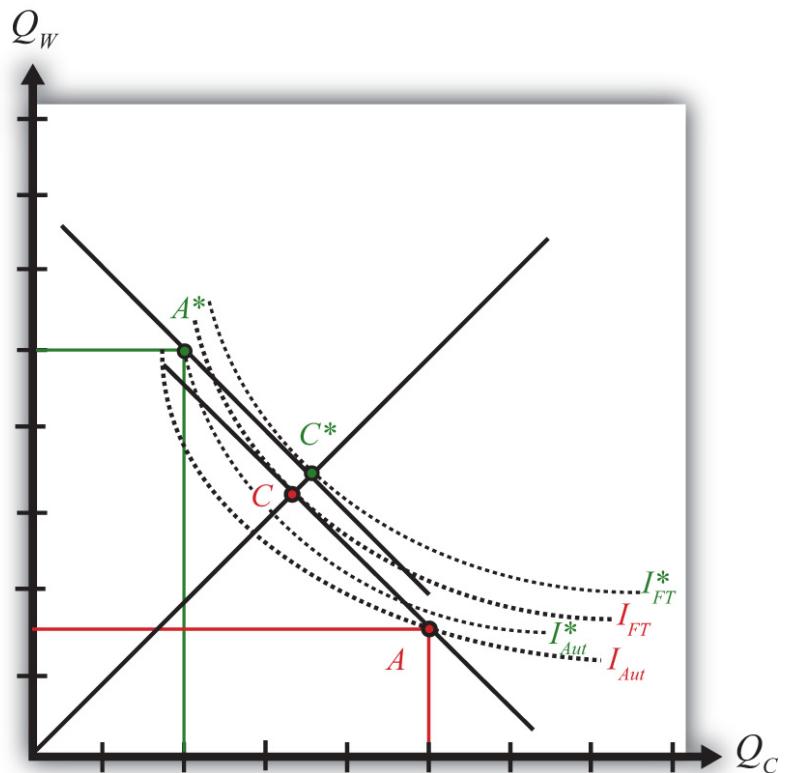
4.11 Aggregate Welfare Effects of Free Trade in the Immobile Factor Model

LEARNING OBJECTIVES

1. Use aggregate indifference curves to demonstrate that a movement to free trade will cause an increase in national welfare in both countries in an immobile factor model.
2. Use national indifference curves to demonstrate the efficiency effects that arise because of free trade in an immobile factor model.

Figure 4.4 "Comparing Free Trade to Autarky" compares autarky and free trade equilibria for the United States and France. The US PPF is given by the red dot at A , while the French PPF is given by the green dot at A^* . We assume both countries share the same aggregate preferences represented by the indifference curves in the diagram.

Figure 4.4 Comparing Free Trade to Autarky



The U.S. autarky production and consumption points are determined where the aggregate indifference curve touches the U.S. PPF at point A. The United States realizes a level of aggregate utility that corresponds to the indifference curve I_{Aut} .

The U.S. production and consumption points in free trade are A and C, respectively. The United States continues to produce at A since factors are immobile between industries but trades to achieve its consumption point at C. In free trade, the United States realizes a level of aggregate utility that corresponds to the indifference curve I_{FT} . Since the free trade indifference curve I_{FT} lies to the northeast of the autarky indifference curve I_{Aut} , national welfare rises as the United States moves to free trade.

France's autarky production and consumption points are determined where the aggregate indifference curve touches France's PPF at point A^* . France realizes a level of aggregate utility that corresponds to the indifference curve I_{Aut}^* .

French production and consumption in free trade occurs at A^* and C^* , respectively. In free trade France realizes a level of aggregate utility that corresponds to the indifference curve I_{FT}^* . Since the free trade indifference curve I_{FT}^* lies to the northeast of the autarky indifference curve I_{Aut}^* , national welfare also rises as France moves to free trade.

This means that free trade will raise aggregate welfare for both countries relative to autarky. Both countries are better off with free trade.

Finally, the aggregate welfare gains from free trade can generally be decomposed into production efficiency gains and consumption efficiency gains. However, since production cannot shift in either country when moving to free trade, there are no production efficiency gains in the immobile factor model. Thus, in the United States, the increase in utility between I_{FT} and I_{Aut} shown in [Figure 4.4 "Comparing Free Trade to Autarky"](#) represents an increase in consumption efficiency only.

KEY TAKEAWAYS

- In an immobile factor model, both countries benefit from free trade because they can both reach a higher aggregate indifference curve.
- In an immobile factor model, there are consumption efficiency improvements but no production efficiency improvements when moving to free trade.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of points A , A^* , C , or C^* in Figure 4.4 "Comparing Free Trade to Autarky", this point provides the highest level of national welfare.
 - b. Of points A , A^* , C , or C^* in Figure 4.4 "Comparing Free Trade to Autarky", this point provides the lowest level of national welfare.
 - c. Of *production efficiency*, *consumption efficiency*, or *both*, improvements in this are shown in the Ricardian model.
 - d. Of *production efficiency*, *consumption efficiency*, or *both*, improvements in this are shown in the immobile factor model.

Chapter 5

The Heckscher-Ohlin (Factor Proportions) Model

The Heckscher-Ohlin (H-O; aka the factor proportions) model is one of the most important models of international trade. It expands upon the Ricardian model largely by introducing a second factor of production. In its two-by-two-by-two variant, meaning two goods, two factors, and two countries, it represents one of the simplest general equilibrium models that allows for interactions across factor markets, goods markets, and national markets simultaneously.

These interactions across markets are one of the important economics lessons displayed in the results of this model. With the H-O model, we learn how changes in supply or demand in one market can feed their way through the factor markets and, with trade, the national markets and influence both goods and factor markets at home and abroad. In other words, all markets are everywhere interconnected.

Among the important results are that international trade can improve economic efficiency but that trade will also cause a redistribution of income between different factors of production. In other words, some will gain from trade, some will lose, but the net effects are still likely to be positive.

The end of the chapter discusses the specific factor model, which represents a cross between the H-O model and the immobile factor model. The implications for income distribution and trade are highlighted.

5.1 Chapter Overview

LEARNING OBJECTIVES

1. Learn the basic assumptions of the Heckscher-Ohlin (H-O) model, especially factor intensity within industries and factor abundancy within countries.
2. Identify the four major theorems in the H-O model.

The factor proportions model was originally developed by two Swedish economists, Eli Heckscher and his student Bertil Ohlin, in the 1920s. Many elaborations of the model were provided by Paul Samuelson after the 1930s, and thus sometimes the model is referred to as the Heckscher-Ohlin-Samuelson (HOS) model. In the 1950s and 1960s, some noteworthy extensions to the model were made by Jaroslav Vanek, and so occasionally the model is called the Heckscher-Ohlin-Vanek model. Here we will simply call all versions of the model either the Heckscher-Ohlin (H-O) model, or simply the more generic “factor proportions model.”

The H-O model incorporates a number of realistic characteristics of production that are left out of the simple Ricardian model. Recall that in the simple Ricardian model only one factor of production, labor, is needed to produce goods and services. The productivity of labor is assumed to vary across countries, which implies a difference in technology between nations. It was the difference in technology that motivated advantageous international trade in the model.

The standard H-O model begins by expanding the number of factors of production from one to two. The model assumes that labor and capital are used in the production of two final goods. Here, capital refers to the physical machines and equipment that are used in production. Thus machine tools, conveyers, trucks, forklifts, computers, office buildings, office supplies, and much more are considered capital.

All productive capital must be owned by someone. In a capitalist economy, most of the physical capital is owned by individuals and businesses. In a socialist economy, productive capital would be owned by the government. In most economies today, the government owns some of the productive capital, but private citizens and businesses own most of the capital. Any person who owns common stock issued by a business has an ownership share in that company and is entitled to dividends or

income based on the profitability of the company. As such, that person is a capitalist—that is, an owner of capital.

The H-O model assumes private ownership of capital. Use of capital in production will generate income for the owner. We will refer to that income as capital “rents.” Thus, whereas the worker earns “wages” for his or her efforts in production, the capital owner earns rents.

The assumption of two productive factors, capital and labor, allows for the introduction of another realistic feature in production: differing factor proportions both across and within industries. When one considers a range of industries in a country, it is easy to convince oneself that the proportion of capital to labor applied in production varies considerably. For example, steel production generally involves large amounts of expensive machines and equipment spread over perhaps hundreds of acres of land, but it also uses relatively few workers. (Note that *relative* here means relative to other industries.) In the tomato industry, in contrast, harvesting requires hundreds of migrant workers to hand-pick and collect each fruit from the vine. The amount of machinery used in this process is relatively small.

In the H-O model, we define the ratio of the quantity of capital to the quantity of labor used in a production process as the **capital-labor ratio**¹. We imagine, and therefore assume, that different industries producing different goods have different capital-labor ratios. It is this ratio (or proportion) of one factor to another that gives the model its generic name: the factor proportions model.

In a model in which each country produces two goods, an assumption must be made as to which industry has the larger capital-labor ratio. Thus if the two goods that a country can produce are steel and clothing and if steel production uses more capital per unit of labor than is used in clothing production, we would say the steel production is **capital intensive**² relative to clothing production. Also, if steel production is capital intensive, then it implies that clothing production must be **labor intensive**³ relative to steel.

1. The ratio of the quantity of capital to the quantity of labor used in a production process.
2. An industry is capital intensive relative to another industry if it has a higher capital-labor ratio in the production process.
3. An industry is labor intensive relative to another industry if it has a higher labor-capital ratio in the production process.

Another realistic characteristic of the world is that countries have different quantities—that is, endowments—of capital and labor available for use in the production process. Thus some countries like the United States are well endowed with physical capital relative to their labor force. In contrast, many less-developed countries have much less physical capital but are well endowed with large labor forces. We use the ratio of the aggregate endowment of capital to the aggregate endowment of labor to define relative factor abundance between countries. Thus if, for example, the United States has a larger ratio of aggregate capital per unit of

labor than France's ratio, we would say that the United States is capital abundant relative to France. By implication, France would have a larger ratio of aggregate labor per unit of capital and thus France would be labor abundant relative to the United States.

The H-O model assumes that the only differences between countries are these variations in the relative endowments of factors of production. It is ultimately shown that (1) trade will occur, (2) trade will be nationally advantageous, and (3) trade will have characterizable effects on prices, wages, and rents when the nations differ in their relative factor endowments and when different industries use factors in different proportions.

It is worth emphasizing here a fundamental distinction between the H-O model and the Ricardian model. Whereas the Ricardian model assumes that production technologies differ between countries, the H-O model assumes that production technologies are the same. The reason for the identical technology assumption in the H-O model is perhaps not so much because it is believed that technologies are really the same, although a case can be made for that. Instead, the assumption is useful in that it enables us to see precisely how differences in resource endowments are sufficient to cause trade and it shows what impacts will arise entirely due to these differences.

The Main Results of the H-O Model

There are four main theorems in the H-O model: the Heckscher-Ohlin (H-O) theorem, the Stolper-Samuelson theorem, the Rybczynski theorem, and the factor-price equalization theorem. The Stolper-Samuelson and Rybczynski theorems describe relationships between variables in the model, while the H-O and factor-price equalization theorems present some of the key results of the model. The application of these theorems also allows us to derive some other important implications of the model. Let us begin with the H-O theorem.

The Heckscher-Ohlin Theorem

The H-O theorem predicts the pattern of trade between countries based on the characteristics of the countries. The H-O theorem says that a capital-abundant country will export the capital-intensive good, while the labor-abundant country will export the labor-intensive good.

4. A country is capital abundant relative to another country if it has a higher capital endowment per labor endowment than the other country.

Here's why. A country that is **capital abundant**⁴ is one that is well endowed with capital relative to the other country. This gives the country a propensity for producing the good that uses relatively more capital in the production

process—that is, the capital-intensive good. As a result, if these two countries were not trading initially—that is, they were in autarky—the price of the capital-intensive good in the capital-abundant country would be bid down (due to its extra supply) relative to the price of the good in the other country. Similarly, in the country that is **labor abundant**⁵, the price of the labor-intensive good would be bid down relative to the price of that good in the capital-abundant country.

Once trade is allowed, profit-seeking firms will move their products to the markets that temporarily have the higher price. Thus the capital-abundant country will export the capital-intensive good since the price will be temporarily higher in the other country. Likewise, the labor-abundant country will export the labor-intensive good. Trade flows will rise until the prices of both goods are equalized in the two markets.

The H-O theorem demonstrates that differences in resource endowments as defined by national abundancies are one reason that international trade may occur.

The Stolper-Samuelson Theorem

The **Stolper-Samuelson theorem**⁶ describes the relationship between changes in output prices (or prices of goods) and changes in factor prices such as wages and rents within the context of the H-O model. The theorem was originally developed to illuminate the issue of how tariffs would affect the incomes of workers and capitalists (i.e., the distribution of income) within a country. However, the theorem is just as useful when applied to trade liberalization.

The theorem states that if the price of the capital-intensive good rises (for whatever reason), then the price of capital—the factor used intensively in that industry—will rise, while the wage rate paid to labor will fall. Thus, if the price of steel were to rise and if steel were capital intensive, the rental rate on capital would rise, while the wage rate would fall. Similarly, if the price of the labor-intensive good were to rise, then the wage rate would rise, while the rental rate would fall.

5. A country is labor abundant relative to another country if it has a higher labor endowment per capital endowment than the other country.

6. A theorem that specifies how changes in output prices affect factor prices in the H-O model. It states that an increase in the price of a good will cause an increase in the price of the factor used intensively in that industry and a decrease in the price of the other factor.

The theorem was later generalized by Ronald Jones, who constructed a magnification effect for prices in the context of the H-O model. The magnification effect allows for analysis of any change in the prices of both goods and provides information about the magnitude of the effects on wages and rents. Most importantly, the magnification effect allows one to analyze the effects of price changes on real wages and real rents earned by workers and capital owners. This is instructive since real returns indicate the purchasing power of wages and rents after accounting for price changes and thus are a better measure of well-being than the wage rate or rental rate alone.

Since prices change in a country when trade liberalization occurs, the magnification effect can be applied to yield an interesting and important result. A movement to free trade will cause the real return of a country's relatively abundant factor to rise, while the real return of the country's relatively scarce factor will fall. Thus if the United States and France are two countries that move to free trade and if the United States is capital abundant (while France is labor abundant), then capital owners in the United States will experience an increase in the purchasing power of their rental income (i.e., they will gain), while workers will experience a decline in the purchasing power of their wage income (i.e., they will lose). Similarly, workers will gain in France, but capital owners will lose.

What's more, the country's abundant factor benefits regardless of the industry in which it is employed. Thus capital owners in the United States would benefit from trade even if their capital is used in the declining import-competing sector. Similarly, workers would lose in the United States even if they are employed in the expanding export sector.

The reasons for this result are somewhat complicated, but the gist can be given fairly easily. When a country moves to free trade, the price of its exported goods will rise, while the price of its imported goods will fall. The higher prices in the export industry will inspire profit-seeking firms to expand production. At the same time, the import-competing industry, suffering from falling prices, will want to reduce production to cut its losses. Thus capital and labor will be laid off in the import-competing sector but will be in demand in the expanding export sector. However, a problem arises in that the export sector is intensive in the country's abundant factor—let's say capital. This means that the export industry wants relatively more capital per worker than the ratio of factors that the import-competing industry is laying off. In the transition there will be an excess demand for capital, which will bid up its price, and an excess supply of labor, which will bid down its price. Hence, the capital owners in both industries experience an increase in their rents, while the workers in both industries experience a decline in their wages.

The Factor-Price Equalization Theorem

The factor-price equalization theorem says that when the prices of the output goods are equalized between countries, as when countries move to free trade, the prices of the factors (capital and labor) will also be equalized between countries. This implies that free trade will equalize the wages of workers and the rents earned on capital throughout the world.

The theorem derives from the assumptions of the model, the most critical of which are the assumptions that the two countries share the same production technology and that markets are perfectly competitive. In a perfectly competitive market, factors are paid on the basis of the value of their marginal productivity, which in turn depends on the output prices of the goods. Thus when prices differ between countries, so will their marginal productivities and hence so will their wages and rents. However, once goods' prices are equalized, as they are in free trade, the value of marginal products is also equalized between countries and hence the countries must also share the same wage rates and rental rates.

Factor-price equalization formed the basis for some arguments often heard in the debates leading up to the approval of the North American Free Trade Agreement (NAFTA) between the United States, Canada, and Mexico. Opponents of NAFTA feared that free trade with Mexico would lower U.S. wages to the level in Mexico. Factor-price equalization is consistent with this fear, although a more likely outcome would be a reduction in U.S. wages coupled with an increase in Mexican wages.

Furthermore, we should note that factor-price equalization is unlikely to apply perfectly in the real world. The H-O model assumes that technology is the same between countries in order to focus on the effects of different factor endowments. If production technologies differ across countries, as we assumed in the Ricardian model, then factor prices would not equalize once goods' prices equalize. As such, a better interpretation of the factor-price equalization theorem applied to real-world settings is that free trade should cause a tendency for factor prices to move together if some of the trade between countries is based on differences in factor endowments.

The Rybczynski Theorem

The **Rybczynski theorem**⁷ demonstrates the relationship between changes in national factor endowments and changes in the outputs of the final goods within the context of the H-O model. Briefly stated, it says that an increase in a country's endowment of a factor will cause an increase in output of the good that uses that factor intensively and a decrease in the output of the other good. In other words, if the United States experiences an increase in capital equipment, then that would cause an increase in output of the capital-intensive good (steel) and a decrease in the output of the labor-intensive good (clothing). The theorem is useful in addressing issues such as investment, population growth and hence labor force growth, immigration, and emigration, all within the context of the H-O model.

7. A theorem that specifies how changes in endowments affect production levels in the H-O model. It states that an increase in a country's endowment of a factor will cause an increase in the output of the good that uses that factor intensively and a decrease in the output of the other good.

The theorem was also generalized by Ronald Jones, who constructed a magnification effect for quantities in the context of the H-O model. The magnification effect allows for analysis of any change in both endowments and provides information about the magnitude of the effects on the outputs of the two goods.

Aggregate Economic Efficiency

The H-O model demonstrates that when countries move to free trade, they will experience an increase in aggregate efficiency. The change in prices will cause a shift in production of both goods in both countries. Each country will produce more of its export good and less of its import good. Unlike the Ricardian model, however, neither country will necessarily specialize in production of its export good. Nevertheless, the production shifts will improve productive efficiency in each country. Also, due to the changes in prices, consumers, in the aggregate, will experience an improvement in consumption efficiency. In other words, national welfare will rise for both countries when they move to free trade.

However, this does not imply that everyone benefits. As the Stolper-Samuelson theorem shows, the model clearly demonstrates that some factor owners will experience an increase in their real incomes, while others will experience a decrease in their factor incomes. Trade will generate winners and losers. The increase in national welfare essentially means that the sum of the gains to the winners will exceed the sum of the losses to the losers. For this reason, economists often apply the compensation principle.

The compensation principle states that as long as the total benefits exceed the total losses in the movement to free trade, then it must be possible to redistribute income from the winners to the losers such that everyone has at least as much as they had before trade liberalization occurred.

Note that the “standard” H-O model refers to the case of two countries, two goods, and two factors of production. The H-O model has been extended to many countries, many goods, and many factors, but most of the exposition in this text, and by economists in general, is in reference to the standard case.

KEY TAKEAWAYS

- The H-O model is a two-country, two-good, two-factor model that assumes production processes differ in their factor intensities, while countries differ in their factor abundances.
- The Rybczynski theorem states there is a positive relationship between changes in a factor endowment and changes in the output of the product that uses that factor intensively.
- The Stolper-Samuelson theorem states there is a positive relationship between changes in a product's price and changes in the payment made to the factor used intensively in that industry.
- The Heckscher-Ohlin theorem predicts the pattern of trade: it says that a capital-abundant (labor-abundant) country will export the capital-intensive (labor-intensive) good and import the labor-intensive (capital-intensive) good.
- The factor-price equalization theorem demonstrates that when product prices are equalized through trade, the factor prices (wages and rents) will be equalized as well.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe the income earned on capital usage.
 - b. The term used to describe the ratio of capital usage to labor usage in an industry.
 - c. The term used to describe an industry that uses more capital per worker than another industry.
 - d. This is by which industries differ from each other in the H-O model.
 - e. This is by which countries differ between each other in the H-O model.
 - f. The name given to the theorem in the H-O model that describes the pattern of trade.
 - g. The name given to the theorem in the H-O model that describes the effects on wages and rents caused by a change in an output price.
 - h. The name given to the theorem in the H-O model that describes the effects on the quantities of the outputs caused by a change in an endowment.
 - i. The name given to the theorem in the H-O model that describes the relationship between factor prices across countries in free trade.

5.2 Heckscher-Ohlin Model Assumptions

LEARNING OBJECTIVE

1. Learn the main assumptions of a two-country, two-good, two-factor Heckscher-Ohlin (or factor proportions) model.

Perfect Competition

Perfect competition in all markets means that the following conditions are assumed to hold.

1. Many firms produce output in each industry such that each firm is too small for its output decisions to affect the market price. This implies that when choosing output to maximize profit, each firm takes the price as given or exogenous.
2. Firms choose output to maximize profit. The rule used by perfectly competitive firms is to choose the output level that equalizes the price (P) with the marginal cost (MC). That is, set $P = MC$.
3. Output is homogeneous across all firms. This means that goods are identical in all their characteristics such that a consumer would find products from different firms indistinguishable. We could also say that goods from different firms are perfect substitutes for all consumers.
4. There is free entry and exit of firms in response to profits. Positive profit sends a signal to the rest of the economy and new firms enter the industry. Negative profit (losses) leads existing firms to exit, one by one, out of the industry. As a result, in the long run economic profit is driven to zero in the industry.
5. Information is perfect. For example, all firms have the necessary information to maximize profit and to identify the positive profit and negative profit industries.

Two Countries

The case of two countries is used to simplify the model analysis. Let one country be the United States, the other France. Note that anything related exclusively to France in the model will be marked with an asterisk.

Two Goods

Two goods are produced by both countries. We assume a barter economy. This means that there is no money used to make transactions. Instead, for trade to occur, goods must be traded for other goods. Thus we need at least two goods in the model. Let the two produced goods be clothing and steel.

Two Factors

Two factors of production, labor and capital, are used to produce clothing and steel. Both labor and capital are homogeneous. Thus there is only one type of labor and one type of capital. The laborers and capital equipment in different industries are exactly the same. We also assume that labor and capital are freely mobile across industries within the country but immobile across countries. Free mobility makes the Heckscher-Ohlin (H-O) model a long-run model.

Factor Constraints

The total amount of labor and capital used in production is limited to the endowment of the country.

The **labor constraint**⁸ is

$$L_C + L_S = L,$$

- 8. A relationship showing that the sum of the labor used in all industries cannot exceed total labor endowment in the economy.
- 9. The total amount of labor resources available to work in an economy during some period of time.
- 10. A relationship showing that the sum of the capital used in all industries cannot exceed total capital endowment in the economy.
- 11. The total amount of capital resources available to work in an economy during some period of time.

where L_C and L_S are the quantities of labor used in clothing and steel production, respectively. L represents the **labor endowment**⁹ of the country. Full employment of labor implies the expression would hold with equality.

The **capital constraint**¹⁰ is

$$K_C + K_S = K,$$

where K_C and K_S are the quantities of capital used in clothing and steel production, respectively. K represents the **capital endowment**¹¹ of the country. Full employment of capital implies the expression would hold with equality.

Endowments

The only difference between countries assumed in the model is a difference in endowments of capital and labor.

Definition

A country is **capital abundant** relative to another country if it has more capital endowment per labor endowment than the other country. Thus in this model the United States is *capital abundant* relative to France if

$$\frac{K}{L} > \frac{K^*}{L^*},$$

where K is the capital endowment and L the labor endowment in the United States and K^* is the capital endowment and L^* the labor endowment in France.

Note that if the United States is capital abundant, then France is labor abundant since the above inequality can be rewritten to get

$$\frac{L^*}{K^*} > \frac{L}{K}.$$

This means that France has more labor per unit of capital for use in production than the United States.

Demand

Factor owners are the consumers of the goods. The factor owners have a well-defined utility function in terms of the two goods. Consumers maximize utility to allocate income between the two goods.

In [Chapter 5 "The Heckscher-Ohlin \(Factor Proportions\) Model"](#), [Section 5.9 "The Heckscher-Ohlin Theorem"](#), we will assume that aggregate preferences can be represented by a homothetic utility function of the form $U = C_S C_C$, where C_S is the amount of steel consumed and C_C is the amount of clothing consumed.

General Equilibrium

The H-O model is a general equilibrium model. The income earned by the factors is used to purchase the two goods. The industries' revenue in turn is used to pay for the factor services. The prices of outputs and factors in an equilibrium are those that equalize supply and demand in all markets simultaneously.

Heckscher-Ohlin Model Assumptions: Production

The production functions in [Table 5.1 "Production of Clothing"](#) and [Table 5.2 "Production of Steel"](#) represent industry production, not firm production. The industry consists of many small firms in light of the assumption of perfect competition.

Table 5.1 Production of Clothing

United States	France
$Q_C = f(L_C, K_C)$	$Q_C^* = f(L_C^*, K_C^*)$

where

Q_C = quantity of clothing produced in the United States, measured in racks

L_C = amount of labor applied to clothing production in the United States, measured in labor hours

K_C = amount of capital applied to clothing production in the United States, measured in capital hours

$f()$ = the clothing production function, which transforms labor and capital inputs into clothing output

* All starred variables are defined in the same way but refer to the production process in France.

Table 5.2 Production of Steel

United States	France
$Q_S = g(L_S, K_S)$	$Q_S^* = g(L_S^*, K_S^*)$

where

Q_S = quantity of steel produced in the United States, measured in tons

L_S = amount of labor applied to steel production in the United States, measured in labor hours

K_S = amount of capital applied to steel production in the United States, measured in capital hours

$g(\cdot)$ = the steel production function, which transforms labor and capital inputs into steel output

*All starred variables are defined in the same way but refer to the production process in France.

Production functions are assumed to be identical across countries within an industry. Thus both the United States and France share the same production function $f(\cdot)$ for clothing and $g(\cdot)$ for steel. This means that the countries share the same technologies. Neither country has a technological advantage over the other. This is different from the Ricardian model, which assumed that technologies were different across countries.

A simple formulation of the production process is possible by defining the unit factor requirements.

Let

$$a_{LC} \left[\frac{\text{labor - hrs}}{\text{rack}} \right]$$

represent the unit labor requirement in clothing production. It is the number of labor hours needed to produce a rack of clothing.

Let

$$a_{KC} \left[\frac{\text{capital - hrs}}{\text{rack}} \right]$$

represent the unit capital requirement in clothing production. It is the number of capital hours needed to produce a rack of clothing.

Similarly,

$$a_{LS} \left[\frac{\text{labor - hrs}}{\text{ton}} \right]$$

is the unit labor requirement in steel production. It is the number of labor hours needed to produce a ton of steel.

And

$$a_{KS} \left[\frac{\text{capital - hrs}}{\text{ton}} \right]$$

is the unit capital requirement in steel production. It is the number of capital hours needed to produce a ton of steel.

By taking the ratios of the unit factor requirements in each industry, we can define a capital-labor (or labor-capital) ratio. These ratios, one for each industry, represent the proportions in which factors are used in the production process. They are also the basis for the model's name.

First, $\frac{a_{KC}}{a_{LC}}$ is the capital-labor ratio in clothing production. It is the proportion in which capital and labor are used to produce clothing.

Similarly, $\frac{a_{KS}}{a_{LS}}$ is the capital-labor ratio in steel production. It is the proportion in which capital and labor are used to produce steel.

Definition

We say that steel production is **capital intensive** relative to clothing production if

$$\frac{a_{KS}}{a_{LS}} > \frac{a_{KC}}{a_{LC}}.$$

This means steel production requires more capital per labor hour than is required in clothing production. Notice that if steel is capital intensive, clothing must be labor intensive.

Clothing production is labor intensive relative to steel production if

$$\frac{a_{LC}}{a_{KC}} > \frac{a_{LS}}{a_{KS}}.$$

This means clothing production requires more labor per capital hour than steel production.

Remember

Factor intensity is a comparison of production processes *across industries* but within a country. Factor abundance is a comparison of endowments *across countries*.

Heckscher-Ohlin Model Assumptions: Fixed versus Variable Proportions

Two different assumptions can be applied in an H-O model: fixed and variable proportions. A fixed proportions assumption means that the capital-labor ratio in each production process is fixed. A variable proportions assumption means that the capital-labor ratio can adjust to changes in the wage rate for labor and the rental rate for capital.

Fixed proportions are more simplistic and also less realistic assumptions. However, many of the primary results of the H-O model can be demonstrated within the context of fixed proportions. Thus the fixed proportions assumption is useful in deriving the fundamental theorems of the H-O model. The variable proportions assumption is more realistic but makes solving the model significantly more difficult analytically. To derive the theorems of the H-O model under variable proportions often requires the use of calculus.

Fixed Factor Proportions

In fixed factor proportions, a_{KC} , a_{LC} , a_{KS} , and a_{LS} are exogenous to the model and are fixed. Since the capital-output and labor-output ratios are fixed, the capital-labor ratios, $\frac{a_{KC}}{a_{LC}}$ and $\frac{a_{KS}}{a_{LS}}$, are also fixed. Thus clothing production must use capital to labor in a particular proportion regardless of the quantity of clothing produced. The ratio of capital to labor used in steel production is also fixed but is assumed to be different from the proportion used in clothing production.

Variable Factor Proportions

Under variable proportions, the capital-labor ratio used in the production process is endogenous. The ratio will vary with changes in the factor prices. Thus if there were a large increase in wage rates paid to labor, producers would reduce their demand for labor and substitute relatively cheaper capital in the production process. This means a_{KC} and a_{LC} are variable rather than fixed. So as the wage and rental rates change, the capital output ratio and the labor output ratio are also going to change.

KEY TAKEAWAYS

- The production process can be simply described by defining unit factor requirements in each industry.
- The capital-labor ratio in an industry is found by taking the ratio of the unit capital and unit labor requirements.
- Factor intensities are defined by comparing capital-labor ratios between industries.
- Factor abundancies are defined by comparing the capital-labor endowment ratios between countries.
- The simple variant of the H-O model assumes the factor proportions are fixed in each industry; a more complex, and realistic, variant assumes factor proportions can vary.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe Argentina if Argentina has more land per unit of capital than Brazil.
 - b. The term used to describe aluminum production when aluminum production requires more energy per unit of capital than steel production.
 - c. The two key terms used in the Heckscher-Ohlin model; one to compare industries, the other to compare countries.
 - d. The term describing the ratio of the unit capital requirement and the unit labor requirement in production of a good.
 - e. The term used to describe when the capital-labor ratio in an industry varies with changes in market wages and rents.
 - f. The assumption in the Heckscher-Ohlin model about unemployment of capital and labor.

5.3 The Production Possibility Frontier (Fixed Proportions)

LEARNING OBJECTIVE

- Plot the labor and capital constraint to derive the production possibility frontier (PPF).

The production possibility frontier (PPF) can be derived in the case of fixed proportions by using the exogenous factor requirements to rewrite the labor and capital constraints. The labor constraint with full employment can be written as

$$a_{LC}Q_C + a_{LS}Q_S = L.$$

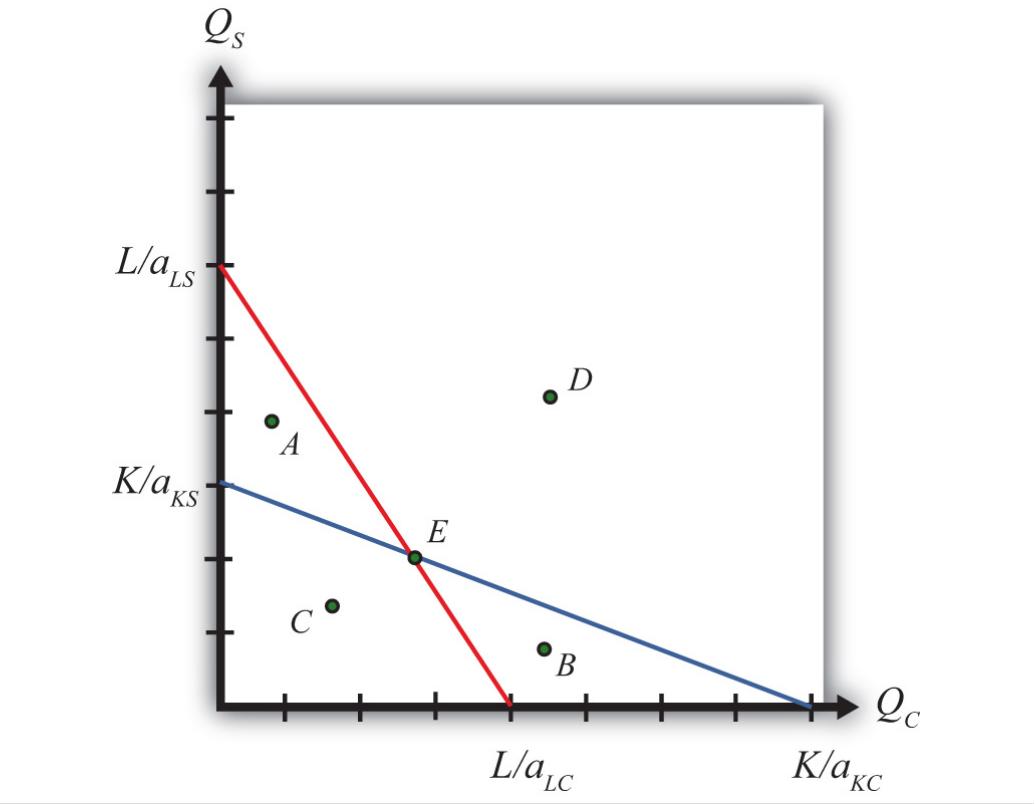
The capital constraint with full employment becomes

$$a_{KC}Q_C + a_{KS}Q_S = K.$$

Each of these constraints contains two endogenous variables: Q_C and Q_S . The remaining variables are exogenous.

We graph the two constraints in [Figure 5.1 "The Labor and Capital Constraints"](#). The red line is the labor constraint. The endpoints $\frac{L}{a_{LC}}$ and $\frac{L}{a_{LS}}$ represent the maximum quantities of clothing and steel that could be produced if all the labor endowments were allocated to clothing and steel production, respectively. All points on the line represent combinations of clothing and steel outputs that could employ all the labor available in the economy. Points outside the constraint, such as B and D, are not feasible production points since there are insufficient labor resources. All points on or within the line, such as A, C, and E, are feasible. The slope of the labor constraint is $-\frac{a_{LC}}{a_{LS}}$.

Figure 5.1 The Labor and Capital Constraints



The blue line is the capital constraint. The endpoints $\frac{K}{a_{KC}}$ and $\frac{K}{a_{KS}}$ represent the maximum quantities of clothing and steel that could be produced if all the capital endowments were allocated to clothing and steel production, respectively. Points on the line represent combinations of clothing and steel production that would employ all the capital in the economy. Points outside the constraint, such as A and D, are not feasible production points since there are insufficient capital resources. Points on or within the line, such as B, C, and E, are feasible. The slope of the capital constraint is $-\frac{a_{KC}}{a_{KS}}$.

The PPF is the set of output combinations that generates full employment of resources—in this case, both labor and capital. Only one point, point E, can simultaneously generate full employment of both labor and capital. Thus point E is the PPF. The production possibility set is the set of all feasible output combinations. The PPS is the area bounded by the axes and the interior section of the labor and capital constraints. Thus at points like A, there is sufficient labor to make production feasible but insufficient capital; thus point A is not a feasible production point. Similarly, at point B there is sufficient capital but not enough labor. Points like C, however, which lie inside (or on) both factor constraints, do represent feasible production points.

Note that the labor constraint is drawn with a steeper slope than the capital constraint. This implies $\frac{a_{LC}}{a_{LS}} > \frac{a_{KC}}{a_{KS}}$, which in turn implies (with cross multiplication) $\frac{a_{KS}}{a_{LS}} > \frac{a_{KC}}{a_{LC}}$. This means that steel is assumed to be capital intensive and clothing production is assumed to be labor intensive. If the slope of the capital constraint had been steeper, then the factor intensities would have been reversed.

KEY TAKEAWAYS

- The PPF in the fixed proportions Heckscher-Ohlin (H-O) model consists of the one point found at the intersection of the linear labor and capital constraints.
- Only those output combinations inside both factor constraint lines are feasible production points within the production possibility set.
- With clothing plotted on the horizontal axis, when the labor constraint is steeper than the capital constraint, clothing is labor intensive.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The description of the PPF in the case of fixed proportions in the Heckscher-Ohlin model.
 - b. The equation for the capital constraint if the unit capital requirement in steel is ten hours per ton, the unit capital requirement in clothing is five hours per rack, and the capital endowment is ten thousand hours.
 - c. The slope of the capital constraint given the information described in Exercise 1b. Include units.
 - d. The equation for the labor constraint if the unit labor requirement in steel is one hour per ton, the unit labor requirement in clothing is three hours per rack, and the labor endowment is one thousand hours.
 - e. The slope of the labor constraint given the information described in Exercise 1d. Include units.
 - f. The capital labor ratio in clothing given the information described in Exercise 1b and Exercise 1d.
 - g. The capital labor ratio in steel given the information described in Exercise 1b and Exercise 1d.

5.4 The Rybczynski Theorem

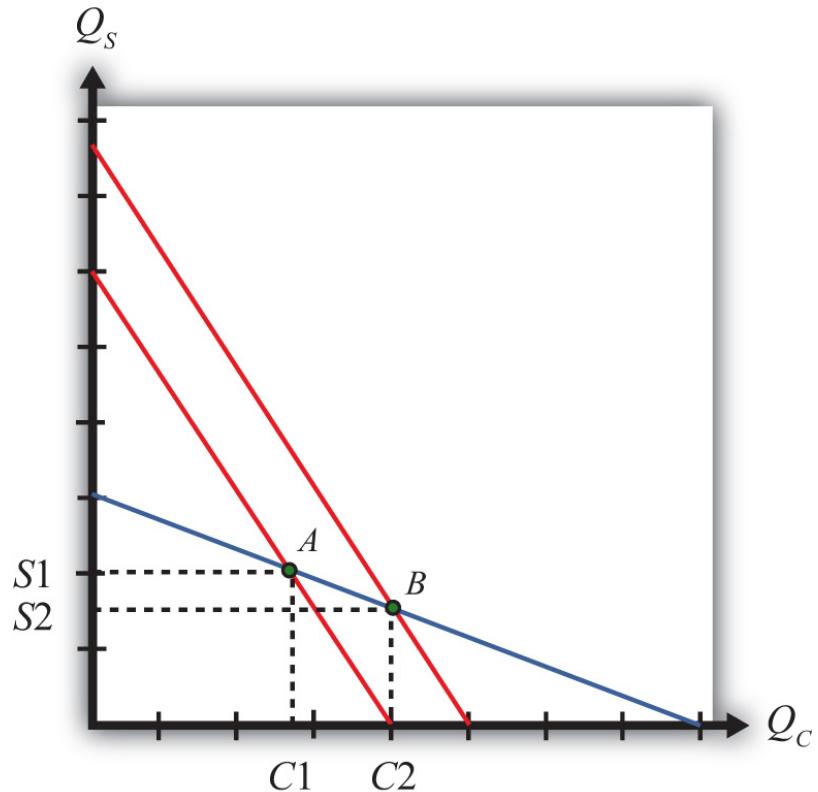
LEARNING OBJECTIVE

1. Use the PPF diagram to show how changes in factor endowments affect production levels at full employment.

The Relationship between Endowments and Outputs

The Rybczynski theorem demonstrates how changes in an endowment affect the outputs of the goods when full employment is maintained. The theorem is useful in analyzing the effects of capital investment, immigration, and emigration within the context of a Heckscher-Ohlin (H-O) model. Consider [Figure 5.2 "Graphical Depiction of Rybczynski Theorem"](#), depicting a labor constraint in red (the steeper lower line) and a capital constraint in blue (the flatter line). Suppose production occurs initially on the PPF at point A.

Figure 5.2 Graphical Depiction of Rybczynski Theorem



Next, suppose there is an increase in the labor endowment. This will cause an outward parallel shift in the labor constraint. The PPF and thus production will shift to point *B*. Production of clothing, the labor-intensive good, will rise from *C*₁ to *C*₂. Production of steel, the capital-intensive good, will fall from *S*₁ to *S*₂.

If the endowment of capital rose, the capital constraint would shift out, causing an increase in steel production and a decrease in clothing production. Recall that since the labor constraint is steeper than the capital constraint, steel is capital intensive and clothing is labor intensive.

This means that, in general, an increase in a country's endowment of a factor will cause an increase in output of the good that uses that factor intensively and a decrease in the output of the other good.

KEY TAKEAWAYS

- The Rybczynski theorem shows there is a positive relationship between changes in a factor endowment and changes in the output of the product that uses that factor intensively.
- The Rybczynski theorem shows there is a negative relationship between changes in a factor endowment and changes in the output of the product that does not use that factor intensively.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the effect on the output of the capital-intensive good caused by a decrease in the labor endowment in a two-factor H-O model.
 - b. Of *increase, decrease, or stay the same*, the effect on the output of the labor-intensive good caused by a decrease in the labor endowment in a two-factor H-O model.
 - c. Of *increase, decrease, or stay the same*, the effect on the output of the capital-intensive good caused by an increase in the capital endowment in a two-factor H-O model.
 - d. Of *increase, decrease, or stay the same*, the effect on the output of the labor-intensive good caused by a decrease in the capital endowment in a two-factor H-O model.
2. Consider an H-O economy in which there are two countries (United States and France), two goods (wine and cheese), and two factors (capital and labor). Suppose an increase in the labor force in the United States causes cheese production to increase. Which factor is used intensively in wine production? Which H-O theorem is applied to get this answer? Explain.

5.5 The Magnification Effect for Quantities

LEARNING OBJECTIVE

1. Learn how the magnification effect for quantities represents a generalization of the Rybczynski theorem by incorporating the relative magnitudes of the changes.

The magnification effect for quantities is a more general version of the Rybczynski theorem. It allows for changes in both endowments simultaneously and allows a comparison of the magnitudes of the changes in endowments and outputs.

The simplest way to derive the magnification effect is with a numerical example.

Suppose the exogenous variables of the model take the values in [Table 5.3 "Numerical Values for Exogenous Variables"](#) for one country.

Table 5.3 Numerical Values for Exogenous Variables

$a_{LC} = 2$	$a_{LS} = 3$	$L = 120$
$a_{KC} = 1$	$a_{KS} = 4$	$K = 120$
where		
L = labor endowment of the country		
K = capital endowment of the country		
a_{LC} = unit labor requirement in clothing production		
a_{KC} = unit capital requirement in clothing production		
a_{LS} = unit labor requirement in steel production		

a_{KS} = unit capital requirement in steel production

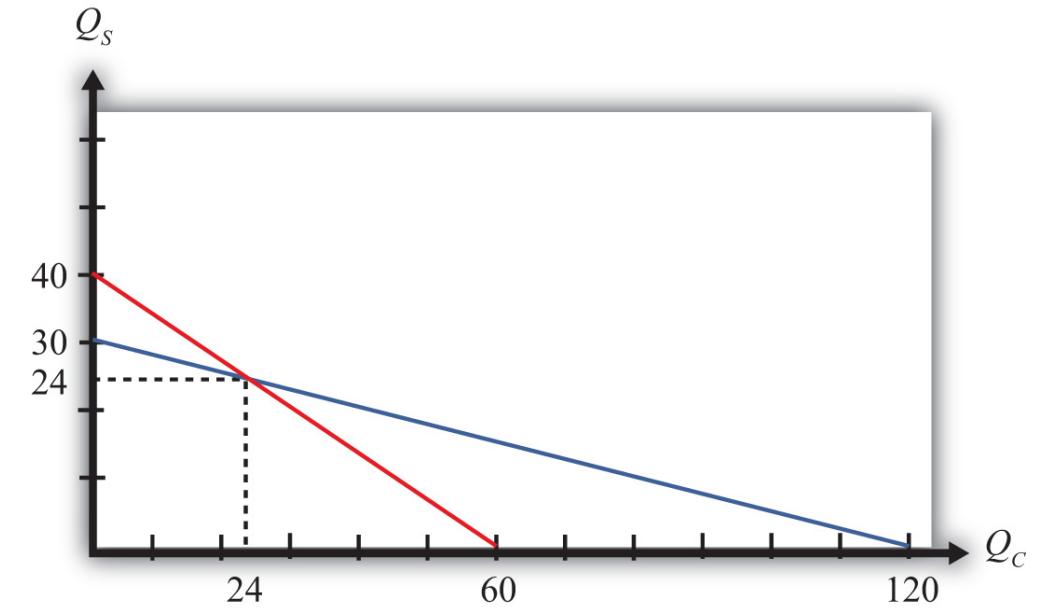
With these numbers, $\frac{a_{KS}}{a_{LS}} \left(\frac{4}{3} \right) > \frac{a_{KC}}{a_{LC}} \left(\frac{1}{2} \right)$ which means that steel production is capital intensive and clothing is labor intensive.

The following are the labor and capital constraints:

- Labor constraint: $2Q_C + 3Q_S = 120$
- Capital constraint: $Q_C + 4Q_S = 120$

We graph these in [Figure 5.3 "Numerical Labor and Capital Constraints"](#). The steeper red line is the labor constraint and the flatter blue line is the capital constraint. The output quantities on the PPF can be found by solving the two constraint equations simultaneously.

Figure 5.3 Numerical Labor and Capital Constraints



A simple method to solve these equations follows.

First, multiply the second equation by (-2) to get

$$2Q_C + 3Q_S = 120$$

and

$$-2Q_C - 8Q_S = -240.$$

Adding these two equations vertically yields

$$0Q_C - 5Q_S = -120,$$

which implies $Q_S = \frac{-120}{-5} = 24$. Plugging this into the first equation above (any equation will do) yields $2Q_C + 3 * 24 = 120$. Simplifying, we get $Q_C = \frac{120 - 72}{2} = 24$. Thus the solutions to the two equations are $Q_C = 24$ and $Q_S = 24$.

Next, suppose the capital endowment, K , increases to 150. This changes the capital constraint but leaves the labor constraint unchanged. The labor and capital constraints now are the following:

- Labor constraint: $2Q_C + 3Q_S = 120$
- Capital constraint: $Q_C + 4Q_S = 150$

Follow the same procedure to solve for the outputs in the new full employment equilibrium.

First, multiply the second equation by (-2) to get

$$2Q_C + 3Q_S = 120$$

and

$$-2Q_C - 8Q_S = -300.$$

Adding these two equations vertically yields

$$0Q_C - 5Q_S = -180,$$

which implies $Q_S = \frac{-180}{-5} = 36$. Plugging this into the first equation above (any equation will do) yields $2Q_C + 3 * 36 = 120$. Simplifying, we get $Q_C = \frac{120 - 108}{2} = 6$. Thus the new solutions are $Q_C = 6$ and $Q_S = 36$.

The Rybczynski theorem says that if the capital endowment rises, it will cause an increase in output of the capital-intensive good (in this case, steel) and a decrease in output of the labor-intensive good (clothing). In this numerical example, Q_S rises from 24 to 36 and Q_C falls from 24 to 6.

Percentage Changes in the Endowments and Outputs

The magnification effect for quantities ranks the percentage changes in endowments and the percentage changes in outputs. We'll denote the percentage change by using a $\hat{}$ above the variable (i.e., \hat{X} = percentage change in X).

Table 5.4 Calculating Percentage Changes in the Endowments and Outputs

$\hat{K} = \frac{150-120}{120} * 100 = +25\%$	The capital stock rises by 25 percent.
$\hat{Q}_S = \frac{36-24}{24} * 100 = +50\%$	The quantity of steel rises by 50 percent.
$\hat{Q}_C = \frac{6-24}{24} * 100 = -75\%$	The quantity of clothing falls by 75 percent.
$\hat{L} = +0\%$	The labor stock is unchanged.

The rank order of the changes in [Table 5.4 "Calculating Percentage Changes in the Endowments and Outputs"](#) is the **magnification effect for quantities**¹²:

$$\hat{Q}_S > \hat{K} > \hat{L} > \hat{Q}_C.$$

The effect is initiated by changes in the endowments. If the endowments change by some percentage, ordered as above, then the quantity of the capital-intensive good (steel) will rise by a *larger* percentage than the capital stock change. The size of the effect is *magnified* relative to the cause.

The quantity of cloth (Q_C) changes by a smaller percentage than the smaller labor endowment change. Its effect is magnified downward.

12. A relationship in the H-O model that specifies the magnitude of output changes in response to changes in the factor endowments.

Although this effect was derived only for the specific numerical values assumed in the example, it is possible to show, using more advanced methods, that the effect will arise for any endowment changes that are made. Thus if the labor endowment

were to rise with no change in the capital endowment, the magnification effect would be

$$\hat{Q}_C > \hat{L} > \hat{K} > \hat{Q}_S.$$

This implies that the quantity of the labor-intensive good (clothing) would rise by a greater percentage than the quantity of labor, while the quantity of steel would fall.

The magnification effect for quantities is a generalization of the Rybczynski theorem. The effect allows for changes in both endowments simultaneously and provides information about the magnitude of the effects. The Rybczynski theorem is one special case of the magnification effect that assumes one of the endowments is held fixed.

Although the magnification effect is shown here under the special assumption of fixed factor proportions and for a particular set of parameter values, the result is much more general. It is possible, using calculus, to show that the effect is valid under any set of parameter values and in a more general variable proportions model.

KEY TAKEAWAYS

- The magnification effect for quantities shows that if the factor endowments change by particular percentages with one greater than the other, then the outputs will change by percentages that are larger than the larger endowment change and smaller than the smaller. It is in this sense that the output changes are magnified relative to the factor changes.
- If the percentage change of the capital endowment exceeds the percentage change of the labor endowment, for example, then output of the good that uses capital intensively will change by a greater percentage than capital changed, while the output of the good that uses labor intensively will change by less than labor changed.

EXERCISES

1. Consider a two-factor (capital and labor), two-good (beer and peanuts) H-O economy. Suppose beer is capital intensive. Let Q_B and Q_P represent the outputs of beer and peanuts, respectively.
 - a. Write the magnification effect for quantities if the labor endowment increases and the capital endowment decreases
 - b. Write the magnification effect for quantities if the capital endowment increases by 10 percent and the labor endowment increases by 5 percent.
 - c. Write the magnification effect for quantities if the labor endowment decreases by 10 percent and the capital endowment decreases by 15 percent.
 - d. Write the magnification effect for quantities if the capital endowment decreases while the labor endowment does not change.
2. Consider a country producing milk and cookies using labor and capital as inputs and described by a Heckscher-Ohlin model. The following table provides outputs for goods and factor endowments before and after a change in the endowments.

TABLE 5.5 OUTPUTS AND ENDOWMENTS

	Initial	After Endowment Change
Milk Output (Q_M)	100 gallons	110 gallons
Cookie Output (Q_C)	100 pounds	80 pounds
Labor Endowment (L)	4,000 hours	4,200 hours
Capital Endowment (K)	1,000 hours	1,000 hours

- a. Calculate and display the magnification effect for quantities in response to the endowment change.
- b. Which product is capital intensive?
- c. Which product is labor intensive?

3. Consider the following data in a Heckscher-Ohlin model with two goods (wine and cheese) and two factors (capital and labor).

$a_{KC} = 5$ hours per pound (unit capital requirement in cheese)

$a_{KW} = 10$ hours per gallon (unit capital requirement in wine)

$a_{LC} = 15$ hours per pound (unit labor requirement in cheese)

$a_{LW} = 20$ hours per gallon (unit labor requirement in wine)

$L = 5,500$ hours (labor endowment)

$K = 2,500$ hours (capital endowment)

- a. Solve for the equilibrium output levels of wine and cheese.
- b. Suppose the labor endowment falls by 100 hours to 5,400 hours. Solve for the new equilibrium output levels of wine and cheese.
- c. Calculate the percentage changes in the outputs and endowments and write the magnification effect for quantities.
- d. Identify which good is labor intensive and which is capital intensive.

5.6 The Stolper-Samuelson Theorem

LEARNING OBJECTIVE

- Plot the zero-profit conditions to show how changes in product prices affect factor prices.

The Stolper-Samuelson theorem demonstrates how changes in output prices affect the prices of the factors when positive production and zero economic profit are maintained in each industry. It is useful in analyzing the effects on factor income either when countries move from autarky to free trade or when tariffs or other government regulations are imposed within the context of a Heckscher-Ohlin (H-O) model.

Due to the assumption of perfect competition in all markets, if production occurs in an industry, then economic profit is driven to zero. The zero-profit conditions in each industry imply

$$P_S = a_{LS} w + a_{KS} r$$

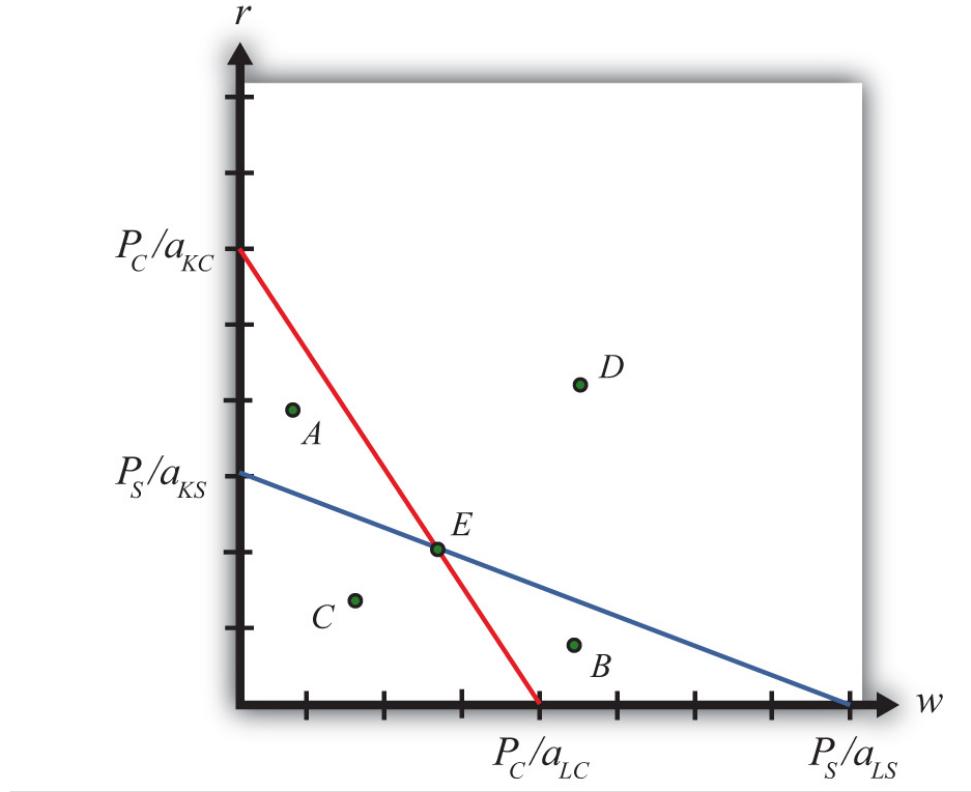
and

$$P_C = a_{LC} w + a_{KC} r,$$

where P_S and P_C are the prices of steel and clothing, respectively; w is the wage paid to labor, and r is the rental rate on capital. Note that $a_{LS} w \left[\frac{\text{labor-hrs}}{\text{ton}} \frac{\$}{\text{labor-hr}} \right] = \frac{\$}{\text{ton}}$ is the dollar payment to workers per ton of steel produced, while $a_{KS} r \left[\frac{\text{capital-hrs}}{\text{ton}} \frac{\$}{\text{capital-hr}} \right] = \frac{\$}{\text{ton}}$ is the dollar payment to capital owners per ton of steel produced. The right-hand-side sum then is the dollars paid to all factors per ton of steel produced. If the payments to factors for each ton produced equal the price per ton, then profit must be zero in the industry. The same logic is used to justify the zero-profit condition in the clothing industry.

We imagine that firms treat prices exogenously since any one firm is too small to affect the price in its market. Because the factor output ratios are also fixed, wages and rentals remain as the two unknowns. In Figure 5.4 "Zero Profit Lines in Clothing and Steel", we plot the two zero-profit conditions in wage-rental space.

Figure 5.4 Zero Profit Lines in Clothing and Steel



The set of all wage and rental rates that will generate zero profit in the steel industry at the price P_S is given by the flatter blue line. At wage and rental combinations above the line, as at points A and D, the per-unit cost of production would exceed the price, and profit would be negative. At wage-rental combinations below the line, as at points B and C, the per-unit cost of production would fall short of the price, and profit would be positive. Notice that the slope of the flatter blue line is $-\frac{P_S/a_{KS}}{P_S/a_{LS}} = -\frac{a_{LS}}{a_{KS}}$

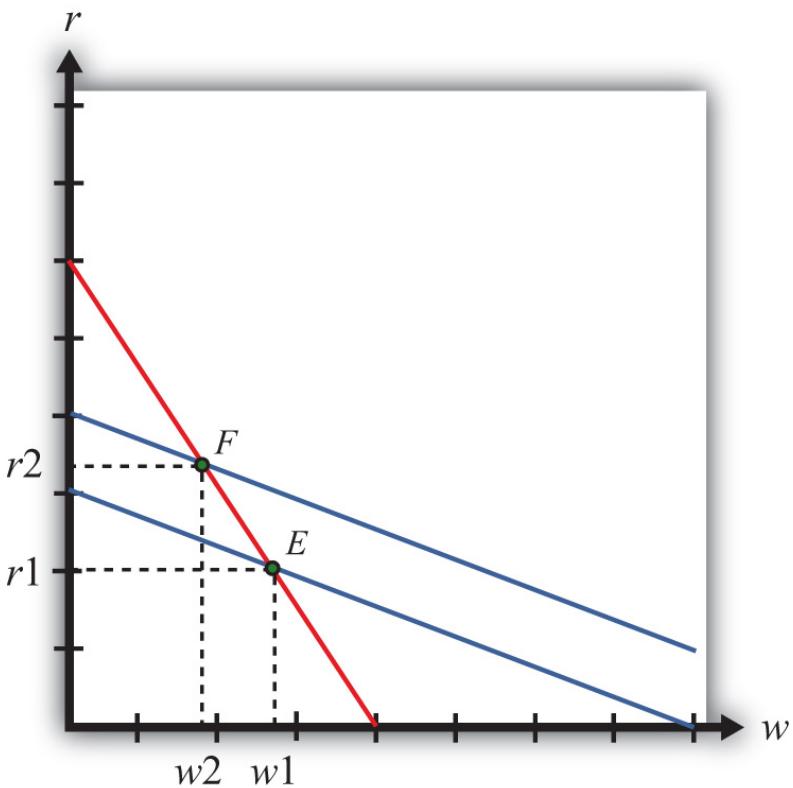
Similarly, the set of all wage-rental rate combinations that will generate zero profit in the clothing industry at price P_C is given by the steeper red line. All wage-rental combinations above the line, as at points B and D, generate negative profit, while wage-rental combinations below the line, as at A and C, generate positive profit. The slope of the steeper red line is $-\frac{P_C/a_{KC}}{P_C/a_{LC}} = -\frac{a_{LC}}{a_{KC}}$.

The only wage-rental combination that can simultaneously support zero profit in both industries is found at the intersection of the two zero-profit lines—point E.

This point represents the equilibrium wage and rental rates that would arise in an H-O model when the price of steel is P_S and the price of clothing is P_C .

Now, suppose there is an increase in the price of one of the goods. Say the price of steel, P_S , rises. This could occur if a country moves from autarky to free trade or if a tariff is placed on imports of steel. The price increase will cause an outward parallel shift in the blue zero-profit line for steel, as shown in [Figure 5.5 "Graphical Depiction of Stolper-Samuelson Theorem"](#). The equilibrium point will shift from E to F , causing an increase in the equilibrium rental rate from r_1 to r_2 and a decrease in the equilibrium wage rate from w_1 to w_2 . Only with a higher rental rate and a lower wage can zero profit be maintained in both industries at the new set of prices. Using the slopes of the zero-profit lines, we can show that $\frac{a_{LC}}{a_{KC}} > \frac{a_{LS}}{a_{KS}}$ which means that clothing is labor intensive and steel is capital intensive. Thus, when the price of steel rises, the payment to the factor used intensively in steel production (capital) rises, while the payment to the other factor (labor) falls.

Figure 5.5 Graphical Depiction of Stolper-Samuelson Theorem



If the price of clothing had risen, the zero-profit line for clothing would have shifted right, causing an increase in the equilibrium wage rate and a decrease in the

rental rate. Thus an increase in the price of clothing causes an increase in the payment to the factor used intensively in clothing production (labor) and a decrease in the payment to the other factor (capital).

This gives us the Stolper-Samuelson theorem: an increase in the price of a good will cause an increase in the price of the factor used intensively in that industry and a decrease in the price of the other factor.

KEY TAKEAWAYS

- The Stolper-Samuelson theorem shows there is a positive relationship between changes in the price of an output and changes in the price of the factor used intensively in producing that product.
- The Stolper-Samuelson theorem shows there is a negative relationship between changes in the price of an output and changes in the price of the factor not used intensively in producing that product.

EXERCISES

1. Consider an H-O economy in which there are two countries (United States and France), two goods (wine and cheese), and two factors (capital and labor). Suppose a decrease in the price of cheese causes a decrease in the wage rate in the U.S. economy. Which factor is used intensively in cheese production in France? Which H-O theorem is used to get this answer? Explain.
2. State what is true about profit in the steel and clothing industry at the wage-rental combination given by the following points in Figure 5.4 "Zero Profit Lines in Clothing and Steel" in the text.
 - a. Point A
 - b. Point B
 - c. Point C
 - d. Point D
 - e. Point E

5.7 The Magnification Effect for Prices

LEARNING OBJECTIVE

1. Learn how the magnification effect for prices represents a generalization of the Stolper-Samuelson theorem by incorporating the relative magnitudes of the changes.

The magnification effect for prices is a more general version of the Stolper-Samuelson theorem. It allows for simultaneous changes in both output prices and compares the magnitudes of the changes in output and factor prices.

The simplest way to derive the magnification effect is with a numerical example.

Suppose the exogenous variables of the model take the values in [Table 5.6 "Numerical Values for Exogenous Variables"](#) for one country.

Table 5.6 Numerical Values for Exogenous Variables

$a_{LS} = 3$	$a_{KS} = 4$	$P_S = 120$
$a_{LC} = 2$	$a_{KC} = 1$	$P_C = 40$
where		
a_{LC} = unit labor requirement in clothing production		
a_{LS} = unit labor requirement in steel production		
a_{KC} = unit capital requirement in clothing production		
a_{KS} = unit capital requirement in steel production		
P_S = the price of steel		

P_C = the price of clothing

With these numbers, $\frac{a_{KS}}{a_{LS}} \left(\frac{4}{3} \right) > \frac{a_{KS}}{a_{LC}} \left(\frac{1}{2} \right)$ which means that steel production is capital intensive and clothing is labor intensive.

The following are the zero-profit conditions in the two industries:

- Zero-profit steel: $3w + 4r = 120$
- Zero-profit clothing: $2w + r = 40$

The equilibrium wage and rental rates can be found by solving the two constraint equations simultaneously.

A simple method to solve these equations follows.

First, multiply the second equation by (-4) to get

$$3w + 4r = 120$$

and

$$-8w - 4r = -160.$$

Adding these two equations vertically yields

$$-5w - 0r = -40,$$

which implies $w = \frac{-40}{-5} = 8$. Plugging this into the first equation above (any equation will do) yields $3*8 + 4r = 120$. Simplifying, we get $r = \frac{120-24}{4} = 24$. Thus the initial equilibrium wage and rental rates are $w = 8$ and $r = 24$.

Next, suppose the price of clothing, P_C , rises from \$40 to \$60 per rack. This changes the zero-profit condition in clothing production but leaves the zero-profit condition in steel unchanged. The zero-profit conditions now are the following:

- Zero-profit steel: $3w + 4r = 120$

- Zero-profit clothing: $2w + r = 60$

Follow the same procedure to solve for the equilibrium wage and rental rates.

First, multiply the second equation by (-4) to get

$$3w + 4r = 120$$

and

$$-8w - 4r = -240.$$

Adding these two equations vertically yields

$$-5w - 0r = -120,$$

which implies $w = \frac{-120}{-5} = 24$. Plugging this into the first equation above (any equation will do) yields $3*24 + 4r = 120$. Simplifying, we get $r = \frac{120-72}{4} = 12$. Thus the new equilibrium wage and rental rates are $w = 24$ and $r = 12$.

The Stolper-Samuelson theorem says that if the price of clothing rises, it will cause an increase in the price paid to the factor used intensively in clothing production (in this case, the wage rate to labor) and a decrease in the price of the other factor (the rental rate on capital). In this numerical example, w rises from \$8 to \$24 per hour and r falls from \$24 to \$12 per hour.

Percentage Changes in the Goods and Factor Prices

The magnification effect for prices ranks the percentage changes in output prices and the percentage changes in factor prices. We'll denote the percentage change by using a $\hat{\Delta}$ above the variable (i.e., $\hat{\Delta}X$ = percentage change in X).

Table 5.7 Calculating Percentage Changes in the Goods and Factor Prices

$\hat{\Delta}P_C = \frac{60-40}{40} * 100 = +50\%$	The price of clothing rises by 50 percent.
$\hat{\Delta}w = \frac{24-8}{8} * 100 = +200\%$	The wage rate rises by 200 percent.

$\hat{r} = \frac{12-24}{24} * 100 = -50\%$	The rental rate falls by 50 percent.
$\hat{P}_S = +0\%$	The price of steel is unchanged.
where	
w = the wage rate	
r = the rental rate	

The rank order of the changes in [Table 5.7 "Calculating Percentage Changes in the Goods and Factor Prices"](#) is the **magnification effect for prices**¹³:

$$\hat{w} > \hat{P}_C > \hat{P}_S > \hat{r}.$$

The effect is initiated by changes in the output prices. These appear in the middle of the inequality. If output prices change by some percentage, ordered as above, then the wage rate paid to labor will rise by a *larger* percentage than the price of steel changes. The size of the effect is *magnified* relative to the cause.

The rental rate changes by a smaller percentage than the price of steel changes. Its effect is magnified downward.

Although this effect was derived only for the specific numerical values assumed in the example, it is possible to show, using more advanced methods, that the effect will arise for any output price changes that are made. Thus if the price of steel were to rise with no change in the price of clothing, the magnification effect would be

$$\hat{r} > \hat{P}_S > \hat{P}_C > \hat{w}.$$

13. A relationship in the H-O model that specifies the magnitude of factor price changes in response to changes in the output prices. It is used to identify the real wage and real rent effects of output price changes.

This implies that the rental rate would rise by a greater percentage than the price of steel, while the wage rate would fall.

The magnification effect for prices is a generalization of the Stolper-Samuelson theorem. The effect allows for changes in both output prices simultaneously and provides information about the magnitude of the effects. The Stolper-Samuelson

theorem is a special case of the magnification effect in which one of the endowments is held fixed.

Although the magnification effect is shown here under the special assumption of fixed factor proportions and for a particular set of parameter values, the result is much more general. It is possible, using calculus, to show that the effect is valid under any set of parameter values and in a more general variable proportions model.

The magnification effect for prices can be used to determine the changes in real wages and real rents whenever prices change in the economy. These changes would occur as a country moves from autarky to free trade and when trade policies are implemented, removed, or modified.

KEY TAKEAWAYS

- The magnification effect for prices shows that if the product prices change by particular percentages with one greater than the other, then the factor prices will change by percentages that are larger than the larger product price change and smaller than the smaller. It is in this sense that the factor price changes are magnified relative to the product price changes.
- If the percentage change in the price of the capital-intensive good exceeds the percentage change in the price of the labor-intensive good, for example, then the rental rate on capital will change by a greater percentage than the price of the capital-intensive good changed, while the wage will change by less than the price of the labor-intensive good.

EXERCISES

1. Consider a country producing milk and cookies using labor and capital as inputs and described by a Heckscher-Ohlin model. The following table provides prices for goods and factors before and after a tariff is eliminated on imports of cookies.

TABLE 5.8 GOODS AND FACTOR PRICES

	Initial (\$)	After Tariff Elimination (\$)
Price of Milk (P_M)	5	6
Price of Cookies (P_C)	10	8
Wage (w)	12	15
Rental rate (r)	20	15

- a. Calculate and display the magnification effect for prices in response to the tariff elimination.
 b. Which product is capital intensive?
 c. Which product is labor intensive?
2. Consider the following data in a Heckscher-Ohlin model with two goods (wine and cheese) and two factors (capital and labor).

$$a_{KC} = 5 \text{ hours per pound (unit capital requirement in cheese)}$$

$$a_{KW} = 10 \text{ hours per gallon (unit capital requirement in wine)}$$

$$a_{LC} = 15 \text{ hours per pound (unit labor requirement in cheese)}$$

$$a_{LW} = 20 \text{ hours per gallon (unit labor requirement in wine)}$$

$$P_C = \$80 \text{ (price of cheese)}$$

$$P_W = \$110 \text{ (price of wine)}$$

- a. Solve for the equilibrium wage and rental rate.

- b. Suppose the price of cheese falls from \$80 to \$75. Solve for the new equilibrium wage and rental rates.
- c. Calculate the percentage changes in the goods prices and factor prices and write the magnification effect for prices.
- d. Identify which good is labor intensive and which is capital intensive.

5.8 The Production Possibility Frontier (Variable Proportions)

LEARNING OBJECTIVE

1. Learn how the shift from a fixed proportions to a variable proportions model affects the presentation of the Heckscher-Ohlin (H-O) model.

The production possibility frontier can be derived in the case of variable proportions by using the same labor and capital constraints used in the case of fixed proportions, but with one important adjustment. Under variable proportions, the unit factor requirements are functions of the wage-rental ratio (w/r). This implies that the capital-labor ratios (which are the ratios of the unit factor requirements) in each industry are also functions of the wage-rental ratio. If there is a change in the equilibrium (for some reason) such that the wage-rental rate rises, then labor will become relatively more expensive compared to capital. Firms would respond to this change by reducing their demand for labor and raising their demand for capital. In other words, firms will substitute capital for labor and the capital-labor ratio will rise in each industry. This adjustment will allow the firm to maintain minimum production costs and thus the highest profit possible. This is the first important distinction between variable and fixed proportions.

The second important distinction is that variable proportions change the shape of the economy's PPF. The labor constraint with full employment can be written as

$$a_{LC}(w/r) Q_C + a_{LS}(w/r) Q_S = L,$$

where a_{LC} and a_{LS} are functions of (w/r) .

The capital constraint with full employment becomes

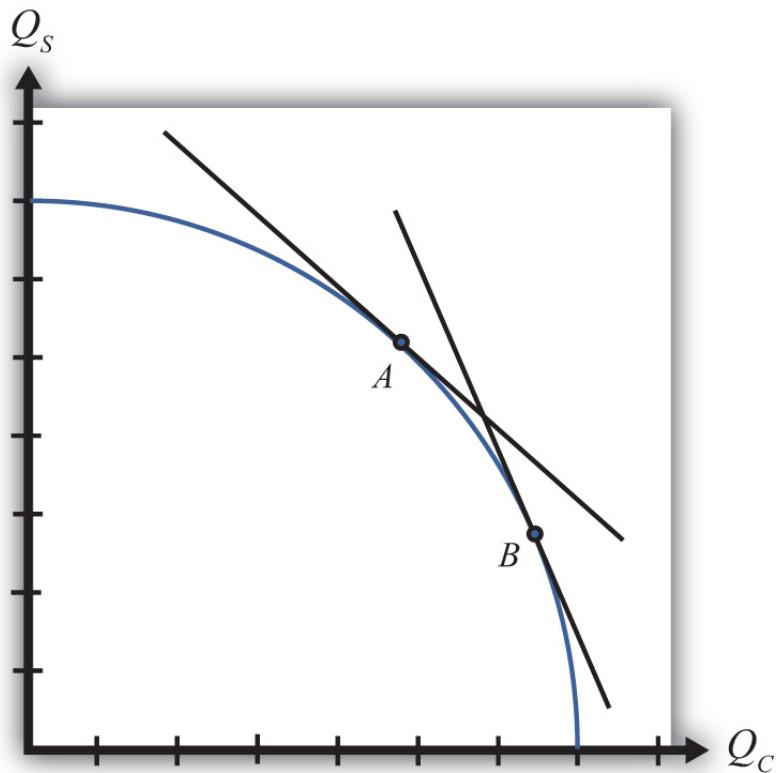
$$a_{KC}(w/r) Q_C + a_{KS}(w/r) Q_S = K,$$

where a_{KC} and a_{KS} are functions of (w/r) .

Under variable proportions, the production possibility frontier takes the traditional bowed-out shape, as shown in [Figure 5.6 "The PPF in the Variable Proportions H-O Model"](#). All points on the PPF will maintain full employment of both labor and capital resources. The slope of a line tangent to the PPF (such as the line through

point A) represents the quantity of steel that must be given up to produce another unit of clothing. As such, the slope of the PPF is the opportunity cost of producing clothing. Since the slope becomes steeper as more and more clothing is produced (as when moving production from point A to B), we say that there is increasing opportunity cost. This means that more steel must be given up to produce one more unit of clothing at point B than at point A in the figure. In contrast, in the Ricardian model the PPF was a straight line that indicated *constant* opportunity costs.

Figure 5.6 The PPF in the Variable Proportions H-O Model



The third important distinction of variable proportions is that the magnification effects, derived previously under a fixed proportions assumption, continue to work under variable proportions. To show this requires a fair amount of advanced math, but a student can rest assured that we can apply the magnification effect even in the more complex variable proportions version of the Heckscher-Ohlin (H-O) model.

KEY TAKEAWAYS

- Variable proportions imply that the capital-labor ratios used in production are varied as wage and rental rates change in the economy.
- Variable proportions imply that the PPF becomes bowed out and continuous, consisting of many output combinations that can be produced with full employment of labor and capital.
- Variable proportions do not invalidate the Rybczynski theorem, the Stolper-Samuelson theorem, or the magnification effects for quantities and prices.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Interpretation given for the slope of the production possibility frontier in the case of variable proportions in the Heckscher-Ohlin model.
 - b. In a variable proportion H-O model, the factor proportions in each industry vary with changes in *these two other variables*.
 - c. Of *increase, decrease, or stay the same*, this is the effect on the capital-labor ratio in an industry when wages fall in a variable proportions H-O model.
 - d. Of *increase, decrease, or stay the same*, this is the effect on the amount of capital used per worker in an industry when rental rates increase in a variable proportions H-O model.
 - e. Of *increase, decrease, or stay the same*, this is the effect on the labor-capital ratio in an industry when wages fall in a variable proportions H-O model.
 - f. Of *increase, decrease, or stay the same*, this is the effect on the capital-labor ratio in the cheese industry when wages increase in a variable proportions H-O model, if cheese is a labor-intensive industry.
 - g. Of *increase, decrease, or stay the same*, this is the effect on the capital-labor ratio in the wine industry when wages increase in a variable proportions H-O model, if wine is a capital-intensive industry.
 - h. Of *increase, decrease, or stay the same*, this is the effect on the capital-labor ratio in an industry when wages fall in a fixed proportions H-O model.

5.9 The Heckscher-Ohlin Theorem

LEARNING OBJECTIVES

1. Learn the Heckscher-Ohlin theorem highlighting the determinants of the pattern of trade.
2. Identify the effects of trade on prices and outputs using a PPF diagram.

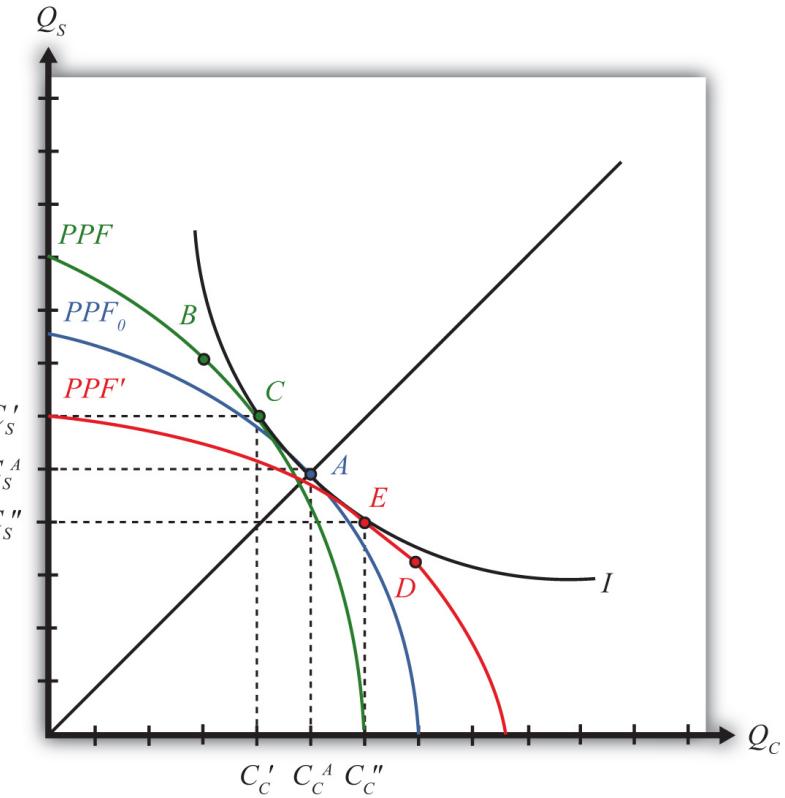
The **Heckscher-Ohlin (H-O) theorem**¹⁴ states that a country that is capital abundant will export the capital-intensive good. Likewise, the country that is labor abundant will export the labor-intensive good. Each country exports that good that it produces relatively better than the other country. In this model, a country's advantage in production arises solely from its relative factor abundance.

The H-O Theorem Graphical Depiction: Variable Proportions

The H-O model assumes that the two countries (United States and France) have identical technologies, meaning they have the same production functions available to produce steel and clothing. The model also assumes that the aggregate preferences are the same across countries. The only difference that exists between the two countries in the model is a difference in resource endowments. We assume that the United States has relatively more capital per worker in the aggregate than does France. This means that the United States is capital abundant compared to France. Similarly, France, by implication, has more workers per unit of capital in the aggregate and thus is labor abundant compared to the United States. We also assume that steel production is capital intensive and clothing production is labor intensive.

14. A theorem that predicts the pattern of trade in the H-O model. It states that the capital-abundant country will export the capital-intensive good and the labor-abundant country will export the labor-intensive good.

Figure 5.7 Endowment Differences and the PPF



The difference in resource endowments is sufficient to generate different PPFs in the two countries such that equilibrium price ratios would differ in autarky. To see why, imagine first that the two countries are identical in every respect. This means they would have the same PPF (depicted as the blue PPF_0 in [Figure 5.7 "Endowment Differences and the PPF"](#)), the same set of aggregate indifference curves, and the same autarky equilibrium. Given the assumption about aggregate preferences—that is, $U = C_C C_S$ —the indifference curve, I , will intersect the countries' PPF at point A , where the absolute value of the slope of the tangent line (not drawn), P_C/P_S , is equal to the slope of the ray from the origin through point A . The slope is given by $\frac{C_S^A}{C_C^A}$. In other words, the autarky price ratio in each country will be given by

$$\left(\frac{P_C}{P_S} \right)_{Aut}^0 = \frac{C_S^A}{C_C^A}.$$

Next, suppose that labor and capital are shifted between the two countries. Suppose labor is moved from the United States to France, while capital is moved from France to the United States. This will have two effects. First, the United States will now

have more capital and less labor, and France will have more labor and less capital than it did initially. This implies that $K/L > K^*/L^*$, or that the United States is capital abundant and France is labor abundant. Second, the two countries' PPFs will shift. To show how, we apply the Rybczynski theorem.

The United States experiences an increase in K and a decrease in L . Both changes will cause an increase in output of the good that uses capital intensively (i.e., steel) and a decrease in output of the other good (clothing). The Rybczynski theorem is derived assuming that output prices remain constant. Thus if prices did remain constant, production would shift from point A to B and the U.S. PPF would shift from the blue PPF_0 to the green PPF in [Figure 5.7 "Endowment Differences and the PPF"](#).

Using the new PPF, we can deduce what the U.S. production point and price ratio would be in autarky given the increase in the capital stock and the decline in the labor stock. Consumption could not occur at point B because first, the slope of the PPF at B is the same as the slope at A because the Rybczynski theorem was used to identify it, and second, homothetic preferences imply that the indifference curve passing through B must have a steeper slope because it lies along a steeper ray from the origin.

Thus to find the autarky production point, we simply find the indifference curve that is tangent to the U.S. PPF. This occurs at point C on the new U.S. PPF along the original indifference curve, I . (Note that the PPF was conveniently shifted so that the same indifference curve could be used. Such an outcome is not necessary but does make the graph less cluttered.) The negative of the slope of the PPF at C is given by the ratio of quantities C_S'/C_C' . Since $C_S'/C_C' > C_S^A/C_C^A$, it follows that the new U.S. price ratio will exceed the one prevailing before the capital and labor shift, that is, $P_C/P_S > (P_C/P_S)^0$. In other words, the autarky price of clothing is higher in the United States after it experiences the inflow of capital and outflow of labor.

France experiences an increase in L and a decrease in K . These changes will cause an increase in output of the labor-intensive good (i.e., clothing) and a decrease in output of the capital-intensive good (steel). If the price were to remain constant, production would shift from point A to D in [Figure 5.7 "Endowment Differences and the PPF"](#), and the French PPF would shift from the blue PPF_0 to the red PPF' .

Using the new PPF, we can deduce the French production point and price ratio in autarky given the increase in the capital stock and the decline in the labor stock. Consumption could not occur at point D since homothetic preferences imply that the indifference curve passing through D must have a flatter slope because it lies

along a flatter ray from the origin. Thus to find the autarky production point, we simply find the indifference curve that is tangent to the French PPF. This occurs at point E on the red French PPF along the original indifference curve, I . (As before, the PPF was conveniently shifted so that the same indifference curve could be used.) The negative of the slope of the PPF at C is given by the ratio of quantities C_S''/C_C'' . Since $C_S''/C_C'' < C_S^A/C_C^A$, it follows that the new French price ratio will be less than the one prevailing before the capital and labor shift—that is, $P_C^*/P_S^* < (P_C/P_S)^0$. This means that the autarky price of clothing is lower in France after it experiences the inflow of labor and outflow of capital.

All of the above implies that as one country becomes labor abundant and the other capital abundant, it causes a deviation in their autarky price ratios. The country with relatively more labor (France) is able to supply relatively more of the labor-intensive good (clothing), which in turn reduces the price of clothing in autarky relative to the price of steel. The United States, with relatively more capital, can now produce more of the capital-intensive good (steel), which lowers its price in autarky relative to clothing. These two effects together imply that

$$\left(\frac{P_C}{P_S} \right)_{Aut}^{US} > \left(\frac{P_C}{P_S} \right)_{Aut}^{FR}.$$

Any difference in autarky prices between the United States and France is sufficient to induce profit-seeking firms to trade. The higher price of clothing in the United States (in terms of steel) will induce firms in France to export clothing to the United States to take advantage of the higher price. The higher price of steel in France (in terms of clothing) will induce U.S. steel firms to export steel to France. Thus the United States, abundant in capital relative to France, exports steel, the capital-intensive good. France, abundant in labor relative to the United States, exports clothing, the labor-intensive good. This is the H-O theorem. Each country exports the good intensive in the country's abundant factor.

KEY TAKEAWAYS

- The H-O theorem states that a country will export that good that is intensive in the country's abundant factor.
- In the standard case, a country will produce more of its export good and less of its import good but will continue to produce both. In other words, specialization does not occur as it does in the Ricardian model.
- Trade is motivated by price differences. A capital-abundant (labor-abundant) country exports the capital-intensive (labor-intensive) good because that product price is initially higher in the labor-abundant (capital-abundant) country.

EXERCISES

1. Consider an H-O economy in which there are two countries (United States and France), two goods (wine and cheese), and two factors (capital and labor). Assume the United States is labor abundant and cheese is labor intensive. What is the pattern of trade in free trade? (State what the United States and France import and export.) Which theorem is applied to get this answer? Explain.

2. Suppose two countries, Malaysia and Thailand, can be described by a variable proportions H-O model. Assume they each produce rice and palm oil using labor and capital as inputs. Suppose Malaysia is capital abundant with respect to Thailand and rice production is labor intensive. Suppose the two countries move from autarky to free trade with each other. In the table below, indicate the effect of free trade on the variables listed in the first column in both Malaysia and Thailand. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 5.9 EFFECTS OF FREE TRADE

	In Malaysia	In Thailand
Price Ratio P_{po}/P_r		
Output of Palm Oil		
Output of Rice		
Exports of Palm Oil		
Imports of Rice		

Chapter 5 The Heckscher-Ohlin (Factor Proportions) Model

	In Malaysia	In Thailand
Capital-Labor Ratio in Palm Oil Production		
Capital-Labor Ratio in Rice Production		

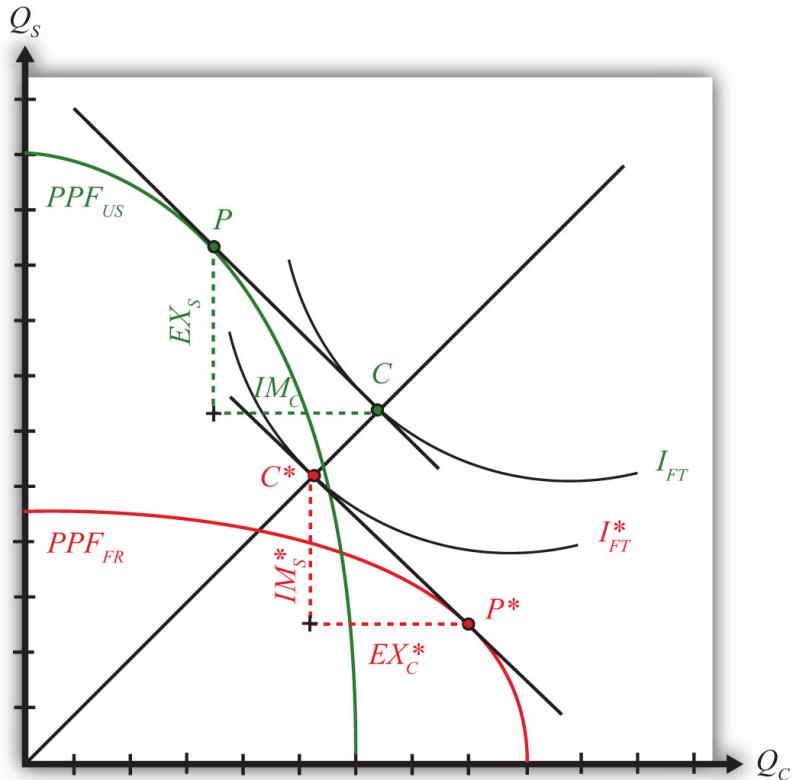
5.10 Depicting a Free Trade Equilibrium in the Heckscher-Ohlin Model

LEARNING OBJECTIVE

1. Learn how to depict a free trade equilibrium on a PPF diagram in the Heckscher-Ohlin (H-O) model.

In [Figure 5.8 "Free Trade Equilibria in an H-O PPF Diagram"](#), we depict free trade equilibria in a Heckscher-Ohlin (H-O) model. The United States is assumed to be capital abundant, which skews its PPF_{US} (in green) in the direction of steel production, the capital-intensive good. France is labor abundant, which skews its PPF_{FR} (in red) in the direction of clothing production, the labor-intensive good. In free trade, each country faces the same price ratio.

Figure 5.8 Free Trade Equilibria in an H-O PPF Diagram



The United States produces at point P . The tangent line at P represents the national income line for the U.S. economy. The equation for the income line is $P_C Q_C + P_S Q_S = NI$, where NI is national income in dollar terms. The slope of the income line is the free trade price ratio $(P_C/P_S)_{FT}$. Consumption in the United States occurs where the aggregate indifference curve I_{FT} , representing preferences, is tangent to the national income line at C . To reach the consumption point, the United States exports EX_S and imports IM_C .

France produces at point P^* . The tangent line at P^* represents the national income line for the French economy. The slope of the income line is also the free trade price ratio $(P_C/P_S)_{FT}$. Consumption in France occurs where the aggregate indifference curve I_{FT}^* , representing preferences, is tangent to the national income line at C^* . Note that since the United States and France are assumed to have the same aggregate *homothetic* preferences and since they face the same price ratio in free trade, consumption for both countries must lie along the same ray from the origin, OC . For France to reach its consumption point, it exports EX_C^* and imports IM_S^* . In order for this to be a free trade equilibrium in a two-country model, U.S. exports of steel must equal French imports of steel ($EX_S = IM_S^*$) and French exports of clothing must equal U.S. imports of clothing ($EX_C^* = IM_C$). In other words, the *U.S. trade triangle* formed by EX_S , IM_C , and the U.S. national income line must be equivalent to *France's trade triangle* formed by EX_C^* , IM_S^* , and the French national income line.

KEY TAKEAWAYS

- The line tangent to the free trade production point on the PPF represents the national income line and has a slope equal to the terms of trade.
- The consumption point in a free trade equilibrium is found as the tangency point of the highest national indifference curve along the national income line tangent to the production point.
- The pattern of trade is shown as the exports and imports needed to move from the production point to the consumption point.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe the slope of the national income line in a two-good, variable proportions H-O model.
 - b. In a two-good, variable proportions H-O model, this occurs where the national income line is tangent to the PPF.
 - c. In a two-good, variable proportions H-O model, this occurs where the national income line is tangent to an indifference curve.
 - d. In a two-good, variable proportions H-O model, these form the base and height of the triangle between the production and consumption points on the PPF diagram.

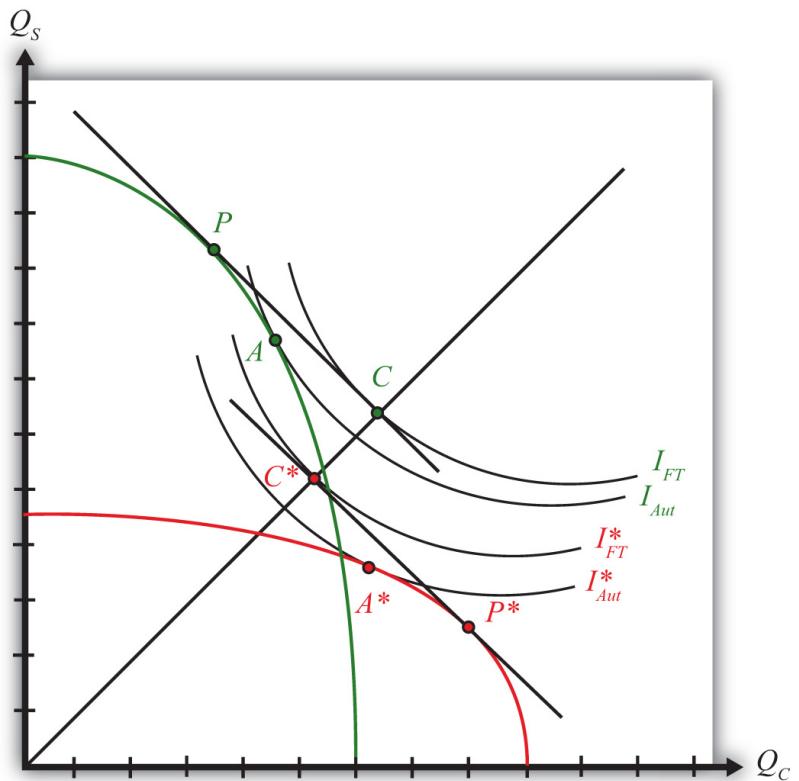
5.11 National Welfare Effects of Free Trade in the Heckscher-Ohlin Model

LEARNING OBJECTIVE

1. Learn how national welfare improvements from free trade can be depicted in a PPF diagram.

Figure 5.9 "National Welfare Effects of Free Trade in the H-O Model" compares autarky and free trade equilibria for the United States and France.

Figure 5.9 *National Welfare Effects of Free Trade in the H-O Model*



The U.S. autarky production and consumption points are determined where the aggregate indifference curve is tangent to the U.S. PPF. This occurs at point A. The United States realizes a level of aggregate utility that corresponds to the indifference curve I_{Aut} .

The U.S. production and consumption points in free trade are P and C , respectively. In free trade, the United States realizes a level of aggregate utility that corresponds to the indifference curve I_{FT} . Since the free trade indifference curve I_{FT} lies to the northeast of the autarky indifference curve I_{Aut} , national welfare rises as the United States moves to free trade.

France's autarky production and consumption points are determined by finding the aggregate indifference curve that is tangent to the French PPF. This occurs at point A^* . France realizes a level of aggregate utility that corresponds to the indifference curve I_{Aut}^* .

French production and consumption points in free trade are P^* and C^* , respectively. In free trade, France realizes a level of aggregate utility that corresponds to the indifference curve I_{FT}^* . Since the free trade indifference curve I_{FT}^* lies to the northeast of the autarky indifference curve I_{Aut}^* , national welfare rises as France moves to free trade.

This means that free trade will raise aggregate welfare for both countries relative to autarky. Both countries are better off with free trade.

However, the use of aggregate indifference curves (or preferences) ignores the issue of income distribution. Although it is correct to conclude from this analysis that both countries benefit from free trade, it is not correct to conclude that all individuals in both countries also benefit from free trade. By calculating changes in real income in the Heckscher-Ohlin (H-O) model, it can be shown that some individuals will likely benefit from free trade, while others will suffer losses. An increase in aggregate welfare means only that the sum of the gains exceeds the sum of the losses.

Another important issue is also typically ignored when using aggregate or national indifference curves to represent a country's preferences. For these curves to make sense, we must assume that income distribution remains the same when moving from one equilibrium to another. That it does not is shown in [Chapter 5 "The Heckscher-Ohlin \(Factor Proportions\) Model"](#), [Section 5.12 "The Distributive Effects of Free Trade in the Heckscher-Ohlin Model"](#). The one way to resolve the issue is to assume that compensation is provided after the redistribution occurs so as to recreate the same income distribution. Compensation is discussed in [Chapter 5 "The Heckscher-Ohlin \(Factor Proportions\) Model"](#), [Section 5.13 "The Compensation Principle"](#).

KEY TAKEAWAY

- In moving from autarky to free trade in an H-O model, both countries can reach a consumption point on a higher national indifference, thereby representing an increase in national welfare.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of points A , A^* , C , C^* , P , or P^* in [Figure 5.9 "National Welfare Effects of Free Trade in the H-O Model"](#), this point provides the highest level of national welfare.
 - b. Of points A , A^* , C , C^* , P , or P^* in [Figure 5.9 "National Welfare Effects of Free Trade in the H-O Model"](#), this point provides the lowest level of national welfare.
 - c. Between indifference curves I_{FT} , I_{FT}^* , I_{AUT} , and I_{AUT}^* in [Figure 5.9 "National Welfare Effects of Free Trade in the H-O Model"](#), points on this curve provide the lowest level of national welfare.
 - d. Between indifference curves I_{FT} , I_{FT}^* , I_{AUT} , and I_{AUT}^* in [Figure 5.9 "National Welfare Effects of Free Trade in the H-O Model"](#), points on this curve provide the highest level of national welfare.
 - e. Of *both increase, both decrease, both stay the same, or one increases and the other decreases*, this is the effect on two countries’ national welfare levels when they move from autarky to free trade in a variable proportions H-O model.
 - f. Of *both increase, both decrease, both stay the same, or one increases and the other decreases*, this is the effect on two countries’ national welfare levels when they move from free trade to autarky in a variable proportions H-O model.

5.12 The Distributive Effects of Free Trade in the Heckscher-Ohlin Model

LEARNING OBJECTIVE

1. Learn how income is redistributed between factors of production when adjusting to free trade.

The term “distributive effects” refers to the distribution of income gains, losses, or both across individuals in the economy. In the Heckscher-Ohlin (H-O) model, there are only two distinct groups of individuals: those who earn their income from labor (workers) and those who earn their income from capital (capitalists). In actuality, many individuals may earn income from both sources. For example, a worker who has deposits in a pension plan that invests in mutual funds has current wage income, but changes in rental rates will affect his or her future capital income. This person’s income stream thus depends on both the return to labor and the return to capital.

For the moment, we shall consider the distributive effects on workers who depend solely on labor income and capitalists who depend solely on capital income. Later we shall consider what happens if individuals receive income from both sources.

To measure gains or losses to workers and capitalists, we must evaluate the effects of free trade on their real incomes. Increases in nominal income are not sufficient to know whether an individual is better off since the price of exportable goods will also rise when a country moves to free trade. By assessing the change in real income, we can determine how the purchasing power of workers and capitalists is affected by the move to free trade.

Suppose there are two countries, the United States and France, producing two goods, clothing and steel, using two factors, capital and labor, according to an H-O model. Suppose steel production is capital intensive and the United States is capital abundant. This implies that clothing production is labor intensive and France is labor abundant.

If these two countries move from autarky to free trade, then, according to the H-O theorem, the United States will export steel to France and France will export clothing to the United States. Also, the price of each country’s export good will rise

relative to each country's import good. Thus in the United States, P_S/P_C rises, while in France P_C/P_S rises.

Next, we apply the magnification effect for prices to each country's price changes.

In the United States, $\left(\frac{P_S}{P_C}\right) \uparrow \Rightarrow \hat{P}_S > \hat{P}_C$ —that is, if the ratio of prices rises, it must mean that the percentage change in P_S is greater than the percentage change in P_C . Then applying the magnification effect for prices implies

$$\hat{r} > \hat{P}_S > \hat{P}_C > \hat{w}.$$

This in turn implies that

$$\frac{r}{P_S} \uparrow, \quad \frac{r}{P_C} \uparrow,$$

which means that the real rent in terms of both steel and clothing rises. And

$$\frac{w}{P_S} \downarrow, \quad \frac{w}{P_C} \downarrow,$$

which means that the real wage in terms of both steel and clothing falls.

Thus individuals in the United States who receive income solely from capital are able to purchase more of each good in free trade relative to autarky. Capitalists are made absolutely better off from free trade. Individuals who receive wage income only are able to purchase less of each good in free trade relative to autarky. Workers are made absolutely worse off from free trade.

In France, $\left(\frac{P_C}{P_S}\right) \uparrow \Rightarrow \hat{P}_C > \hat{P}_S$ —that is, the percentage change in P_C is greater than the percentage change in P_S . Then, according to the magnification effect for prices,

$$\hat{w} > \hat{P}_C > \hat{P}_S > \hat{r}.$$

This in turn implies that

$$\frac{w}{P_C} \uparrow, \frac{w}{P_S} \uparrow,$$

which means that the real wage in terms of both clothing and steel rises. And

$$\frac{r}{P_C} \downarrow, \frac{r}{P_S} \downarrow,$$

which means that the real rent in terms of both clothing and steel falls.

Thus individuals in France who receive wage income only are able to purchase more of each good in free trade relative to autarky. Workers are made absolutely better off from free trade. Individuals in France who receive income solely from capital are able to purchase less of each good in free trade relative to autarky. Capitalists are made absolutely worse off from free trade.

These results imply that both countries will experience a redistribution of income when moving from autarky to free trade. Some individuals will gain from trade, while others will lose. Distinguishing the winners and losers more generally can be done by referring to the fundamental basis for trade in the model. Trade occurs because of differences in endowments between countries. The United States is assumed to be capital abundant, and when free trade occurs, capitalists in the United States benefit. France is assumed to be labor abundant, and when free trade occurs, workers in France benefit. Thus, in the H-O model, a country's relatively abundant factor gains from trade, while a country's relatively scarce factor loses from trade.

It is worth noting that the redistribution of income is between factors of production and not between industries. The H-O model assumes that workers and capital are homogenous and are costlessly mobile between industries. This implies that all workers in the economy receive the same wage and all capital receives the same rent. Thus if workers benefit from trade in the H-O model, it means that all workers in both industries benefit. In contrast to the immobile factor model, one need not be affiliated with the export industry in order to benefit from trade. Similarly, if capital loses from trade, then capitalists suffer losses in both industries. One need not be affiliated with the import industry to suffer losses.

KEY TAKEAWAYS

- In the H-O model, when countries implement free trade, output prices, wages, and rents on capital change.
- If a country is abundant in capital (labor), then a movement to free trade will increase real rents (wages) and decrease real wages (rents). In other words, income is redistributed from workers (capital owners) to capital owners (workers).
- Because labor and capital are assumed to be homogeneous factors, workers (capital owners) in both industries realize identical real income effects.
- The redistribution of income in the H-O model is based on which factor an individual owns, not on which industry an individual works in (as it is in the immobile factor model).

EXERCISES

1. Consider an H-O economy in which there are two countries (United States and France), two goods (wine and cheese), and two factors (capital and labor).
 - a. Suppose France exports wine, the capital-intensive good. Which factor benefits from free trade in the United States? Explain.
 - b. Suppose workers in France benefit when tariffs are increased on cheese imports. Which factor is used intensively in cheese production? What is France's abundant factor? Explain.
2. Suppose two countries, Malaysia and Thailand, can be described by a variable proportions H-O model. Assume they each produce rice and palm oil using labor and capital as inputs. Suppose Malaysia is capital abundant with respect to Thailand and rice production is labor intensive. Suppose the two countries move from autarky to free trade with each other. In the table below, indicate the effect of free trade on the variables listed in the first column in both Malaysia and Thailand. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 5.10 EFFECTS OF FREE TRADE

	In Malaysia	In Thailand
Price Ratio P_{po}/P_r		

Chapter 5 The Heckscher-Ohlin (Factor Proportions) Model

	In Malaysia	In Thailand
Real Wage in Terms of Palm Oil		
Real Wage in Terms of Rice		
Real Rental Rate in Terms of Palm Oil		
Real Rental Rate in Terms of Rice		

5.13 The Compensation Principle

LEARNING OBJECTIVES

1. Learn how compensation, consisting of a redistribution of income after a new equilibrium is reached, can support an equal distribution of benefits arising from free trade.
2. Learn why economists suggest *lump-sum* redistributions as the most effective way to compensate the losers with gains from the winners.

The Heckscher-Ohlin model generates several important conclusions for a country that moves from autarky to free trade:

- Aggregate national welfare rises—this is displayed as achieving a higher level of utility on a set of national indifference curves.
- Income is redistributed among individuals within the economy—this is shown by applying the magnification effect for prices to the price changes that arise in moving from autarky to free trade. It is shown that the real income of a country's relatively abundant factor rises while the real income of a country's relatively scarce factor falls.

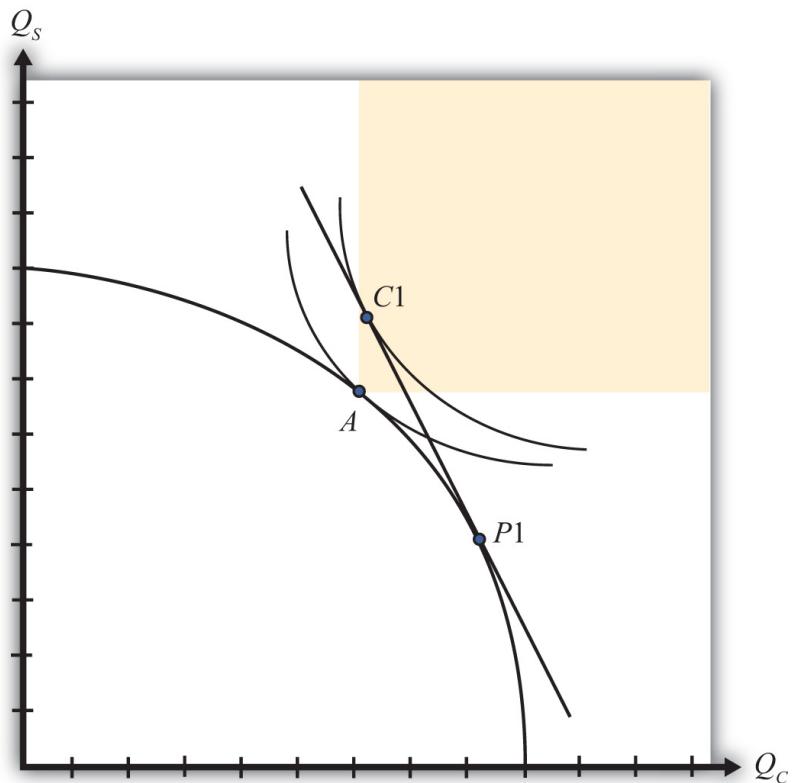
A reasonable question at this juncture, then, is whether the gains to some individuals exceed the losses to others and, if so, whether it is possible to redistribute income to ensure that everyone is absolutely better off with trade than he or she was in autarky. In other words, is it possible for the winners from free trade to *compensate* the losers in such a way that everyone is left better off than he or she was in autarky?

The answer to this is yes in most circumstances. The primary reason is that the move to free trade improves production and consumption efficiency, which can make it possible for the country to consume more of both goods with trade compared to autarky.

Consider [Figure 5.10 "Compensation in the H-O Model"](#). Point A on the PPF represents the autarky production and consumption point for this economy. The shaded region represents the set of consumption points that provides at least as much of one good and more of the other relative to the autarky equilibrium. Suppose that in free trade production moves to P1 and consumption moves to C1.

Since C_1 lies within the shaded region, the country consumes more clothing and more steel in the aggregate than it had consumed in autarky. However, in moving from autarky to free trade, some factors have experienced increases in income, while others have suffered losses. This means that some individuals consume less of both goods in free trade, while others consume more of both goods.

Figure 5.10 Compensation in the H-O Model



However, since there are more of both goods in the aggregate, it is conceivable that government intervention, which takes some of the extra goods away from the winners, could sufficiently compensate the losers and leave everyone better off in trade.

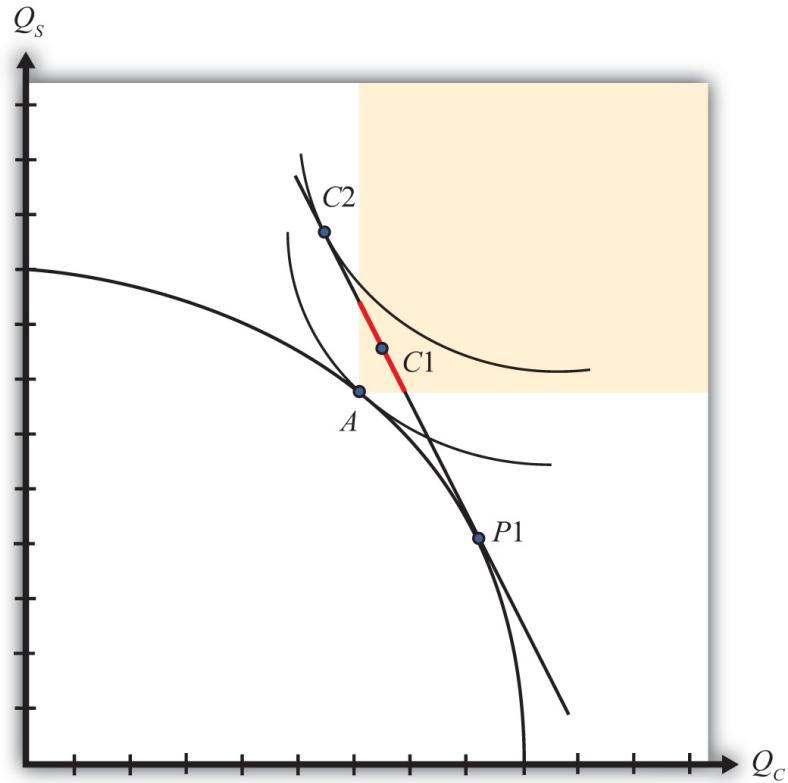
The possibility of an effective redistribution depends in some circumstances on the way in which the redistribution is implemented. For example, taxes and subsidies could redistribute income from winners to losers but would simultaneously affect the domestic prices of the goods, which would affect consumption decisions and so on. With the secondary effects of taxes and subsidies, it becomes uncertain whether a redistribution policy would work. For this reason, economists will often talk about making a *lump-sum* redistribution or transfer. Lump-sum transfers are analogous to the transfers from rich to poor made by the infamous character Robin Hood.

Essentially, goods must be stolen away from the winners, after they have made their consumption choices, and given to the losers, also after they have made their consumption choices. Furthermore, the winners and losers must not know or expect that a redistribution will be made, lest that knowledge affect their consumption choices beforehand. Thus a lump-sum redistribution is exactly what Robin Hood achieves. He steals from the wealthy, after they've purchased their goods, and gives to the poor, who were not expecting such a gift.

Although lump-sum compensations make perfect sense in theory, or in principle, it is worth noting how impractical they are. There is no government that has tried to institutionalize this process by creating a Division of Robin Hoodian Transfers. In practice, lump-sum transfers rarely occur.

Compensation may not always be as straightforward as in the previous example, however. Another possible outcome in a free trade equilibrium is for more of one good to be consumed but less of another relative to autarky. In other words, the free trade consumption point may occur at a point like C_2 in [Figure 5.11 "Compensation Difficulties"](#). In this case, it would not be possible to compensate everyone with as much steel as they had in autarky since the economy is consuming less steel in the free trade equilibrium. However, even in this case it is potentially possible to arrange a redistribution scheme. The reason is that the economy could potentially choose a consumption point along the red line segment, as at point C_1 . Since the red segment lies in the range in which more of both goods is available, compensation to make everyone better off with trade remains a possibility.

Figure 5.11 Compensation Difficulties



Thus it is always possible to find a free trade consumption point and an appropriate lump-sum compensation scheme such that everyone is at least as well off with trade as they had been in autarky.

KEY TAKEAWAYS

- Because the sum of the benefits accruing to the winners exceeds the sum of the losses to the losers from free trade, it is possible to conceive of an income redistribution, or compensation, scheme that will assure that all individuals gain from trade.
- To avoid upsetting the optimal decisions made by producers and consumers in a free trade equilibrium, the most effective compensation scheme involves lump-sum transfers from winners to losers.
- Lump-sum transfers, although effective in theory, are virtually impossible to implement in practice.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a policy response that can alleviate the losses caused to some groups and assure that everyone gains from trade liberalization.
 - b. Of points A, C1, or P1 in [Figure 5.10 "Compensation in the H-O Model"](#), this point provides the highest level of national welfare.
 - c. Of points A, C1, or P1 in [Figure 5.10 "Compensation in the H-O Model"](#), this point provides the lowest level of national welfare.
 - d. A type of compensation reminiscent of Robin Hood.
 - e. Lump-sum transfers were conceived as a way to avoid the effects of taxes or subsidies on these decisions.
2. When a country moves to free trade, there are several ways to identify an improvement in the nation’s welfare. One method requires information about the nation’s preferences, especially the trade-offs between consumption of different goods; the other method does not. Explain.

5.14 Factor-Price Equalization

LEARNING OBJECTIVE

1. Understand the relationship between wages and rents across countries in the Heckscher-Ohlin (H-O) model.

The fourth major theorem that arises out of the Heckscher-Ohlin (H-O) model is called the factor-price equalization theorem. Simply stated, the theorem says that when the prices of the output goods are equalized between countries as they move to free trade, then the prices of the factors (capital and labor) will also be equalized between countries. This implies that free trade will equalize the wages of workers and the rents earned on capital throughout the world.

The theorem derives from the assumptions of the model, the most critical of which is the assumption that the two countries share the same production technology and that markets are perfectly competitive.

In a perfectly competitive market, the return to a factor of production depends on the value of its marginal productivity. The marginal productivity of a factor, like labor, in turn depends on the amount of labor being used as well as the amount of capital. As the amount of labor rises in an industry, labor's marginal productivity falls. As the amount of capital rises, labor's marginal productivity rises. Finally, the value of productivity depends on the output price commanded by the good in the market.

In autarky, the two countries face different prices for the output goods. The difference in prices alone is sufficient to cause a deviation in wages and rents between countries because it affects the marginal productivity. However, in addition, in a variable proportions model the difference in wages and rents also affects the capital-labor ratios in each industry, which in turn affects the marginal products. All of this means that for various reasons the wage and rental rates will differ between countries in autarky.

Once free trade is allowed in outputs, output prices will become equal in the two countries. Since the two countries share the same marginal productivity relationships, it follows that only one set of wage and rental rates can satisfy these

relationships for a given set of output prices. Thus free trade will equalize goods' prices and wage and rental rates.

Since the two countries face the same wage and rental rates, they will also produce each good using the same capital-labor ratio. However, because the countries continue to have different quantities of factor endowments, they will produce different quantities of the two goods.

This result contrasts with the Ricardian model. In that model, production technologies are assumed to be different in the two countries. As a result, when countries move to free trade, real wages remain different from each other; the country with higher productivities will have higher real wages.

In the real world, it is difficult to know whether production technologies are different, similar, or identical. Supporting identical production technology, one could argue that state-of-the-art capital can be moved anywhere in the world. On the other hand, one might counter by saying that just because the equipment is the same doesn't mean the workforces will operate the equipment similarly. There will likely always remain differences in organizational abilities, workforce habits, and motivations.

One way to apply these model results to the real world might be to say that to the extent that countries share identical production capabilities, there will be a tendency for factor prices to converge as freer trade is realized.

KEY TAKEAWAYS

- The factor-price equalization theorem says that when the product prices are equalized between countries as they move to free trade in the H-O model, then the prices of the factors (capital and labor) will also be equalized between countries.
- Factor-price equalization arises largely because of the assumption that the two countries have the same technology in production.
- Factor-price equalization in the H-O model contrasts with the Ricardian model result in which countries could have different factor prices after opening to free trade.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. *This key technology assumption* assures that factor-price equalization will occur in free trade in an H-O model.
 - b. The factor price equalization theorem says these will be equalized between countries if goods prices become equalized because of trade.
 - c. The factor price equalization theorem says these will be equalized between countries if factor prices become equalized because of factor migration.
2. Suppose there are two countries, Japan and the Philippines, described by a variable proportions H-O model. Suppose they produce two goods, rice and chicken, using two factors, labor and capital. Let rice be capital intensive and the Philippines be labor abundant.
 - a. If these are the only two countries and if they do not trade, explain how the price of rice and chicken will differ between the two countries.
 - b. If these are the only two countries and if they do not trade, explain how the wages and rental rates on capital will differ between the two countries.
 - c. When trade opens between the countries what happens to the price of rice and chicken in the Philippines?
 - d. When trade opens between the countries what happens to the wages and rents in the Philippines?
 - e. When trade opens between the countries what happens to the wages and rents in Japan?
 - f. When trade is free between the two countries, how do the wages and rents compare between the two countries?
3. Suppose there are two countries, Japan and the Philippines, as described in Exercise 2 above. Suppose goods trade is restricted between the countries and that factor mobility between countries suddenly becomes free.

- a. Describe the pattern of factor flows that would occur between the two countries and explain why these flows occur.
- b. Describe the effect of the factor flows on the wages and rents in the two countries.
- c. Apply the magnification effect for quantities to explain how the outputs of rice and chicken will change in Japan and the Philippines.
- d. After factor flows reach a new equilibrium, explain how goods' prices will differ between the two countries.

5.15 The Specific Factor Model: Overview

LEARNING OBJECTIVE

1. Learn the basic assumptions and results of the specific factor (SF) model.

The specific factor (SF) model was originally discussed by Jacob Viner, and it is a variant of the Ricardian model. Hence the model is sometimes referred to as the Ricardo-Viner model. The model was later developed and formalized mathematically by Ronald Jones (1971)^{See R. W. Jones, "A Three-Factor Model in Theory, Trade and History," in *Trade, Balance of Payments and Growth*, ed. J. N. Bhagwati, R. W. Jones, R. A. Mundell, and J. Vanek (Amsterdam: North-Holland Publishing Co., 1971).} and Michael Mussa (1974)^{Michael Mussa, "Tariffs and the Distribution of Income: The Importance of Factor Specificity, Substitutability, and Intensity in the Short and Long-Run," *Journal of Political Economy*, 82, no. 6 (1974): 1191–1203..} Jones referred to it as the two-good, three-factor model. Mussa developed a simple graphical depiction of the equilibrium that can be used to portray some of the model's results. It is this view that is presented in most textbooks.

The model's name refers to its distinguishing feature—that one factor of production is assumed to be “specific” to a particular industry. A *specific* factor is one that is stuck in an industry or is immobile between industries in response to changes in market conditions. A factor may be immobile between industries for a number of reasons. Some factors may be specifically designed (in the case of capital) or specifically trained (in the case of labor) for use in a particular production process. In these cases, it may be impossible, or at least difficult or costly, to move these factors across industries. See [Chapter 4 "Factor Mobility and Income Redistribution"](#), [Section 4.2 "Domestic Factor Mobility"](#) and [Chapter 4 "Factor Mobility and Income Redistribution"](#), [Section 4.3 "Time and Factor Mobility"](#) for more detailed reasons for factor immobility.

The SF model is designed to demonstrate the effects of trade in an economy in which one factor of production is specific to an industry. The most interesting results pertain to the changes in the distribution of income that would arise as a country moves to free trade.

Basic Assumptions

The SF model assumes that an economy produces two goods using two factors of production, capital and labor, in a perfectly competitive market. One of the two factors of production, typically capital, is assumed to be specific to a particular industry—that is, it is completely immobile. The second factor, labor, is assumed to be freely and costlessly mobile between the two industries. Because capital is immobile, one could assume that capital in the two industries is different, or differentiated, and thus is not substitutable in production. Under this interpretation, it makes sense to imagine that there are really three factors of production: labor, specific capital in Industry 1, and specific capital in Industry 2.

These assumptions place the SF model squarely between an immobile factor model and the Heckscher-Ohlin (H-O) model. In an immobile factor model, all the factors of production are specific to an industry and cannot be moved. In an H-O model, both factors are assumed to be freely mobile—that is, neither factor is specific to an industry. Since the mobility of factors in response to any economic change is likely to increase over time, we can interpret the immobile factor model results as short-run effects, the SF model results as medium-run effects, and the H-O model results as long-run effects.

Production of Good 1 requires the input of labor and capital specific to Industry 1. Production of Good 2 requires labor and capital specific to Industry 2. There is a fixed endowment of sector-specific capital in each industry as well as a fixed endowment of labor. Full employment of labor is assumed, which implies that the sum of the labor used in each industry equals the labor endowment. Full employment of sector-specific capital is also assumed; however, in this case the sum of the capital used in all the *firms* within the industry must equal the endowment of sector-specific capital.

The model assumes that firms choose an output level to maximize profit, taking prices and wages as given. The equilibrium condition will have firms choosing an output level, and hence a labor usage level, such that the market-determined wage is equal to the value of the marginal product of the last unit of labor. The **value of the marginal product**¹⁵ is the increment of revenue that a firm will obtain by adding another unit of labor to its production process. It is found as the product of the price of the good in the market and the marginal product of labor. Production is assumed to display diminishing returns because the fixed stock of capital means that each additional worker has less capital to work with in production. This means that each additional unit of labor will add a smaller increment to output, and since the output price is fixed, the value of the marginal product declines as labor usage rises. When all firms behave in this way, the allocation of labor between the two industries is uniquely determined.

15. The increment in revenue that a firm will obtain by adding another unit of labor to its production process.

The production possibility frontier (PPF) will exhibit increasing opportunity costs. This is because expansion of one industry is possible by transferring labor out of the other industry, which must therefore contract. Due to the diminishing returns to labor, each additional unit of labor switched will have a smaller effect on the expanding industry and a larger effect on the contracting industry. This means that the graph of the PPF in the SF model will look similar to the PPF in the variable proportion H-O model. However, in relation to a model in which both factors were freely mobile, the SF model PPF will lie everywhere inside the H-O model PPF. This is because the lack of mobility of one factor inhibits firms from taking full advantage of efficiency improvements that would be possible if both factors can be freely reallocated.

Specific Factor Model Results

The SF model is used to demonstrate the effects of economic changes on labor allocation, output levels, and factor returns. Many types of economic changes can be considered, including a movement to free trade, the implementation of a tariff or quota, growth of the labor or capital endowment, or technological changes. This section will focus on effects that result from a change in prices. In an international trade context, prices might change when a country liberalizes trade or when it puts into place additional barriers to trade.

When the model is placed into an international trade context, differences of some sort between countries are needed to induce trade. The standard approach is to assume that countries differ in the amounts of the specific factors used in each industry relative to the total amount of labor. This would be sufficient to cause the PPFs in the two countries to differ and could potentially generate trade. Under this assumption, the SF model is a simple variant of the H-O model. However, the results of the model are not sensitive to this assumption. Trade may arise due to differences in endowments, differences in technology, differences in demands, or some combination. The results derive as long as there is a price change, for whatever reason.

So suppose, in a two-good SF model, that the price of one good rises. If the price change is the result of trade liberalization, then the industry whose price rises is in the export sector. The price increase would set off the following series of adjustments. First, higher export prices would initially raise profits in the export sector since wages and rents may take time to adjust. The value of the marginal product in exports would rise above the current wage, and that would induce the firms to hire more workers and expand output. However, to induce the movement of labor, the export firms would have to raise the wage that they pay. Since all labor is alike (the model assumes labor is homogeneous), the import-competing sector would have to raise its wages in step so as not to lose all of its workers. The higher

wages would induce the expansion of output in the export sector (the sector whose price rises) and a reduction in output in the import-competing sector. The adjustment would continue until the wage rises to a level that equalizes the value of the marginal product in both industries.

The return to capital in response to the price change would vary across industries. In the import-competing industry, lower revenues and higher wages would combine to reduce the return to capital in that sector. However, in the export sector, greater output and higher prices would combine to raise the return to capital in that sector.

The real effects of the price change on wages and rents are somewhat more difficult to explain but are decidedly more important. Remember that absolute increases in the wage, or the rental rate on capital, does not guarantee that the recipient of that income is better off, since the price of one of the goods is also rising. Thus the more relevant variables to consider are the real returns to capital (real rents) in each industry and the real return to labor (real wages).

Ronald Jones (1971) derived a magnification effect for prices in the SF model that demonstrated the effects on the real returns to capital and labor in response to changes in output prices. In the case of an increase in the price of an export good and a decrease in the price of an import good, as when a country moves to free trade, the magnification effect predicts the following impacts:

1. The real return to capital in the export industry will rise with respect to purchases of both exports and imports.
2. The real return to capital in the import-competing industry will fall with respect to purchases of both exports and imports.
3. The real wage to workers in both industries will rise with respect to purchases of the import good and will fall with respect to purchases of the export good.

This result means that when a factor of production, like capital, is immobile between industries, a movement to free trade will cause a redistribution of income. Some individuals—owners of capital in the export industry—will benefit from free trade. Other individuals—owners of capital in the import-competing industries—will lose from free trade. Workers, who are freely mobile between industries, may gain or may lose since the real wage in terms of exports rises while the real wage in terms of imports falls. If workers' preferences vary, then those individuals who have a relatively high demand for the export good will suffer a welfare loss, while those individuals who have a relatively strong demand for imports will experience a welfare gain.

Notice that the clear winners and losers in this model are distinguishable by industry. As in the immobile factor model, the factor specific to the export industry benefits, while the factor specific to the import-competing industry loses.

KEY TAKEAWAYS

- The specific factor (SF) model is designed to evaluate the real-world phenomenon that some factors of production are more mobile between industries than others. It does that by assuming that one factor (capital) cannot move between industries, while the other factor (labor) can freely move.
- In all other respects, the SF model is like the H-O model.
- The SF model shows that upon opening to free trade, the real rents in the exports industry rise, real rents in the import-competing industry fall, and real wages in both industries may rise or fall.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used in economic models to describe a factor of production that is so specialized that it can only be used in a single industry.
 - b. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade on the real return to specific capital in the export industry.
 - c. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade on the real return to specific capital in the import industry.
 - d. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade on the real wages when labor is the mobile factor in a specific factor model.
 - e. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade on the real wage with respect to the imported good when labor is the mobile factor in a specific factor model.
 - f. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade on the real wage with respect to the exported good when labor is the mobile factor in a specific factor model.

5.16 The Specific Factor Model

LEARNING OBJECTIVES

1. Learn the detailed assumptions of the specific factor model.
2. Learn how price changes affect wages, rents, and factor returns using the Mussa diagram.
3. Learn the real wage and real rent effects of free trade in a specific factor model.

Consider an economy with two perfectly competitive industries, textiles and steel. Suppose the output of both products requires labor and capital as factor inputs. However, we'll imagine the capital used in textile production consists of equipment such as looms, while the capital used in steel production requires equipment such as blast furnaces. Since each type of capital is designed for use in a specific production process, we call it "specific capital." We can imagine that if the capital from one industry were shifted to another, its productivity in the new industry would be zero. Simply imagine the usefulness of a blast furnace in textile production and you should see the point! Thus for capital to remain fully employed, it must remain in the same industry—it is immobile, or stuck in its respective industry.

We assume labor, on the other hand, is homogenous and perfectly freely mobile between the two industries. This will imply that a firm's choice problem is reduced to the decision of how much labor to hire and how much to produce to maximize its profits, given that it has a fixed amount of capital available to use. We'll assume for simplicity that the capital stock in each industry is exogenously fixed and there is no investment in new capital.

Single-Firm Equilibrium in the Specific Factor Model

In this context, a firm will maximize its profits when it produces a level of output such that the wage it must pay to workers is equal to the value of the marginal product at the chosen level of output. This is written in equation form for a textile firm as follows:

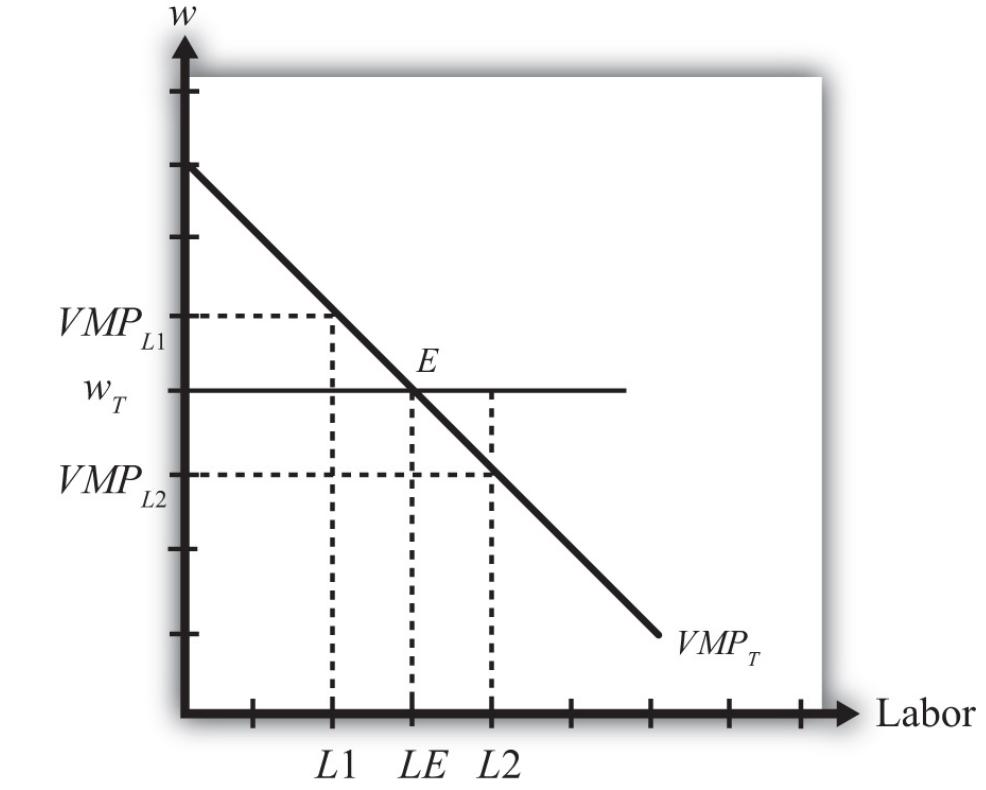
$$w = P_T MP_T.$$

The left-hand side of the equation represents the hourly wage the firm pays its workers. The right-hand side is the value of the marginal product, which consists of the product of the market price of output (P_T) and the marginal product of production (MP_T). The marginal product, in turn, represents the additional output that can be obtained by increasing the labor input by one unit. For example, if $MP_T = 10$, this means that by adding one more hour of labor, ten additional meters of cloth could be produced. The units of the expression MP_T are meters of cloth per hour of labor (m/hr.). When multiplied by the price, measured as dollars per meter, the product, $P_T MP_T$, yields *the number of dollars that could be earned per hour of additional labor applied in production*. This then is the definition of the value of the marginal product in this context. It is measured in dollars per hour, the same as the wage is measured—a good thing since they must be equal to each other!

To see why this condition will hold when the firm maximizes profit, we will graph these expressions in [Figure 5.12 "Specific Factor Model—Single-Firm Equilibrium"](#), which depicts the value of a marginal product line for a representative textile firm, $VMP_T = P_T MP_T$, and the market wage rate, w_T , with respect to the labor supply.

The wage is assumed to be exogenous to each firm and is independent of the labor supply. Hence it is drawn as a horizontal line at the level of the wage, w_T . Later the wage will be determined endogenously through the interaction of the two industries. Nevertheless, firms in both industries recognize they are too small to influence the market wage and make decisions based on an exogenously given wage.

Figure 5.12 Specific Factor Model—Single-Firm Equilibrium



The value of the marginal product is a decreasing function of labor. This means that at higher levels of labor usage, each additional unit of labor applied to production adds fewer units of output. The intuition for this is straightforward. Imagine more and more workers being assigned to use the same machine in a production process. Each additional worker may help in the production process and add output (thus $MP > 0$), but as more and more are added, overcrowding will set in and each person will find less and less to do that is helpful. Thus the marginal product will fall. Since we draw the value of the marginal product line under the assumption that there is a fixed amount of specific capital in the industry, the same overcrowding argument applies at the larger industry scale.

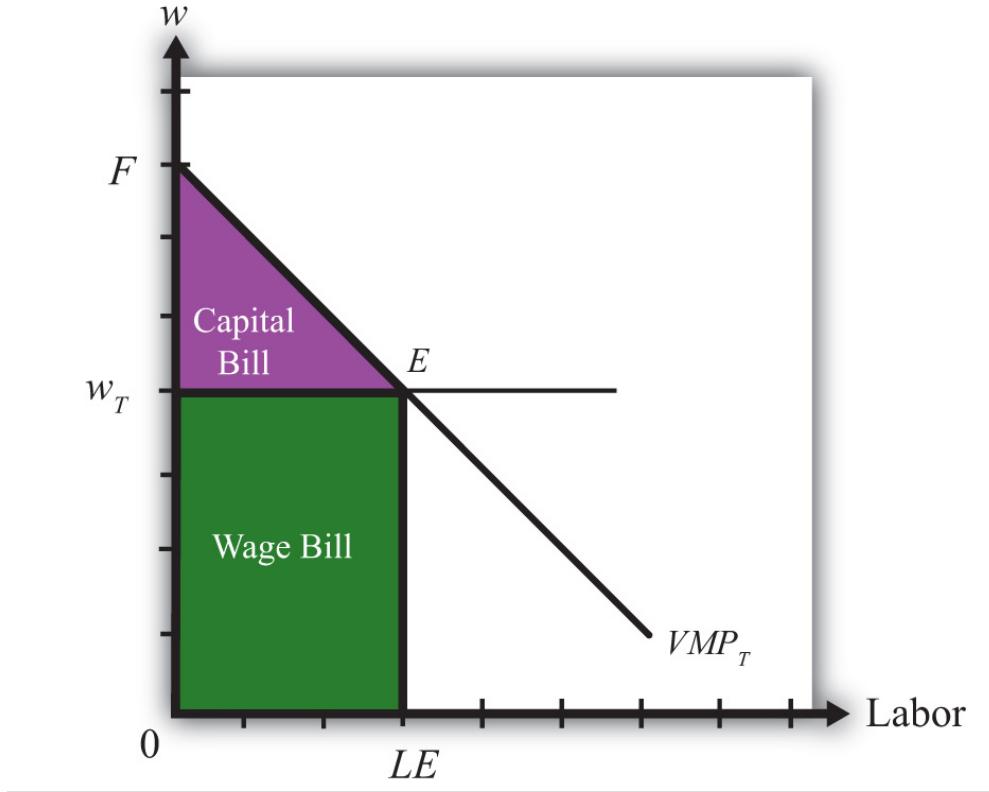
The position of the VMP line is dependent on the market price and the amount of specific capital, both assumed to be exogenous. If the price of the product rises (falls), the VMP line shifts upward (downward). The same applies for changes in the amount of specific capital. If the amount of specific capital in the industry were to rise (fall), the VMP line would shift upward (downward).

The profit-maximizing choice of labor input by the industry is determined at level LE on the horizontal axis, where the wage w_T is equal to the value of the marginal product VMP_T at point E . To see why, consider what it would mean if the industry chose a different labor input, say L_1 . At L_1 , $VMP_L^1 > w_T$. This says that the additional revenue earned by expanding labor input by one unit exceeds the additional cost of adding one more unit of labor. Thus adding one more unit of labor must raise profit, which means that L_1 cannot be the profit-maximizing choice—it must lie to the right of L_1 . Next consider labor input L_2 . At L_2 , $VMP_L^2 < w_T$. This says that the additional revenue earned by expanding labor input by one unit is less than the cost of adding one more unit of labor. Thus adding one more unit of labor must lower profit, which means that L_2 cannot be the profit-maximizing choice—it must lie to the left of L_2 . Finally, consider labor input LE . At LE , $VMP_{LE} = w_T$. This says that the additional revenue earned by expanding labor input by one unit equals the additional cost of adding one more unit of labor. Thus adding one more unit of labor has no effect on profit, which means that LE must be the profit-maximizing choice.

Factor Payments

In [Figure 5.13 "Specific Factor Model—Factor Payments"](#), we consider ways to represent the factor payments made in an equilibrium. Consider a wage rate w_T and an equilibrium labor input given by LE . The product of these two, $w_T LE$, represents the total amount of money that must be paid to workers in the industry and is referred to as the wage bill. It is the charges incurred by the owners (i.e., the bill that must be paid) to hire the workers. It is represented by the green shaded area.

Figure 5.13 Specific Factor Model—Factor Payments



The total amount of revenue earned by the firm on the market is given by the total shaded area (green + purple). This corresponds to the area under the VMP_T line between 0 and LE units of labor. Without the use of calculus, it is difficult to describe why this is so. Nonetheless, since the VMP gives the additional revenue earned for each additional unit of labor, one can imagine beginning back at $L = 0$ and increasing labor in small increments. The vertical distance to the VMP line would be added to the total revenue for every increment in labor. Adding each of these vertical lines together between $L = 0$ and $L = L^E$ yields total revenue earned by the firm and is given by the total shaded area.

Finally, since there are only two factors of production—labor and specific capital—it must follow that the total revenue equals the sum of the wage bill and the capital bill, where the capital bill represents the total amount of money paid to the capital owners. In equation form we could write

$$\text{total revenue} = \text{wage bill} + \text{capital bill}.$$

Since the total revenue is given by the total shaded area and the wage bill is given by the lower shaded area, the capital bill must be given by the upper purple shaded

area. Again, this area represents the total amount of money the firm must pay to the owners of capital used in production. It is not the rental rate, however. The rental rate is given by the rental bill divided by the total quantity of capital units used in production. In other words, the rental rate in textiles, r_T , is given by

$$r_T = \text{rental bill}/K_T,$$

where K_T is the fixed amount of specific capital available for use in the industry.

Similarly, the wage rate in textiles, w_T , is given by

$$w_T = \text{wage bill}/L_E.$$

Two-Firm Equilibrium in the Specific Factor Model

The economy consists of two industries, textiles and steel, each of which is choosing labor input so as to maximize profit. Thus when both industries operate and both maximize profit,

$$w_T = VMP_T$$

for textiles and

$$w_S = VMP_S$$

for steel, where w_T and w_S are the wage rates paid to workers in textiles and steel, respectively. With homogeneous and perfectly mobile labor, another condition must also hold, namely, the labor constraint:

$$L_T + L_S = L.$$

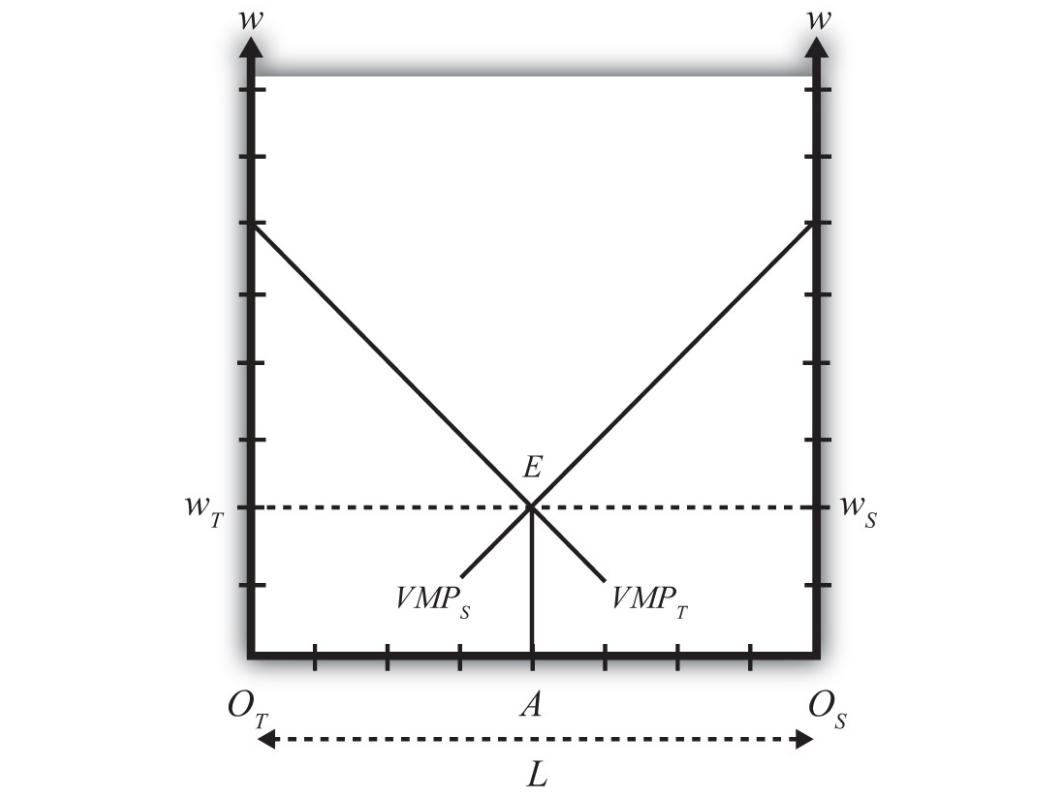
In other words, the labor used in textile production (L_T) plus the labor used in steel production (L_S) must equal the total labor endowment available in the economy (L). Finally, because labor is homogeneous and perfectly mobile between industries, wages must be equalized in equilibrium between the two industries. Thus

$$w_T = w_S.$$

All four conditions must be satisfied simultaneously in an equilibrium in this model. To represent this equilibrium and to provide a medium to analyze potential

changes, we present a diagram developed by Mussa (1974). The diagram (shown in [Figure 5.14 "Specific Factor Model—Mussa Diagram"](#)) is unique in that it presents all four conditions together on the same graph. The horizontal axis of the diagram plots the labor supply. The vertical axis plots the wage and the value of the marginal products.

Figure 5.14 Specific Factor Model—Mussa Diagram



The horizontal length of [Figure 5.14 "Specific Factor Model—Mussa Diagram"](#), $O_T O_S$, represents the labor endowment (L), the total amount of labor available for use in the economy. The VMP_T line slopes down from the left as presented before.

However, the VMP_S line slopes down from the right. This is because the point O_S corresponds to zero units of labor used in steel production and $O_T O_S$ units of labor used in textiles. As we move to the left from O_S , labor used in steel increases, while labor used in textiles decreases. Thus the VMP_S line is flipped and drawn with respect to its origin at O_S . Every point along the horizontal axis corresponds to an allocation of labor between the two industries satisfying the labor constraint condition. Thus at a point like A , $O_T A$ units of labor are used in textile production (L_T) and $O_S A$ units of labor are used in steel production (L_S). The sum of the two equals $O_T O_S$, which is the total labor endowment (L).

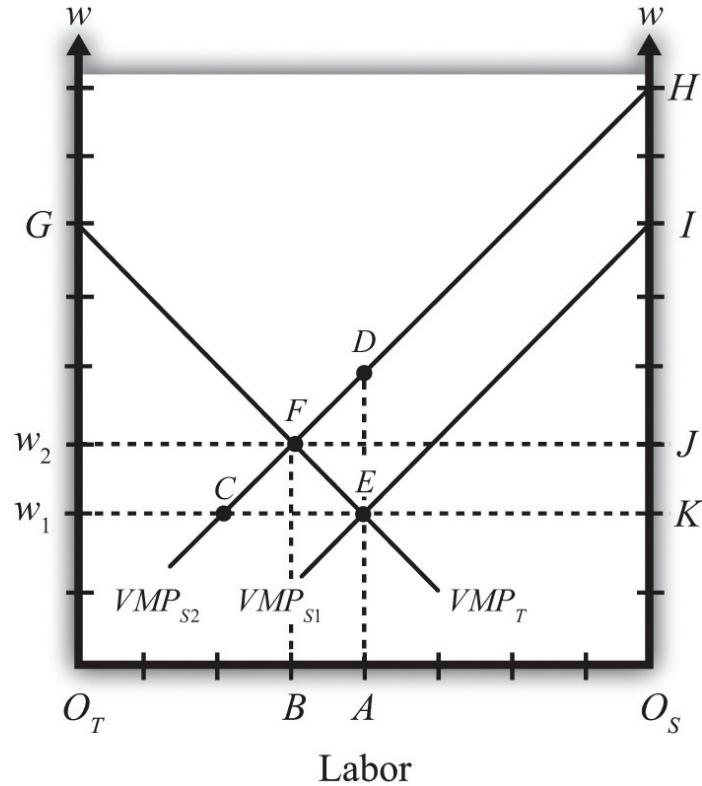
At point E in [Figure 5.14 "Specific Factor Model—Mussa Diagram"](#), the two VMP lines intersect so that $VMP_T = VMP_S$, determining the unique wage rate $w = w_T = w_S$ using all the available labor, O_TO_S . Thus at point E all four equilibrium conditions listed are satisfied.

Effects of a Price Increase

Prices will change whenever a country moves from autarky to free trade or when a country imposes a trade or domestic policy. At this stage, we will simply consider the effects of a price change within the context of the model without specifying why the change occurred. (In more technical terms, we say the price change is exogenous.) Later, we'll introduce several situations to see how trade or trade policies will affect outcomes in the specific factor (SF) model.

Suppose we begin with a country producing textiles and steel in an initial equilibrium given by point E in [Figure 5.15 "Effects of a Price Increase"](#). The original value of the marginal product lines is given by VMP_{T1} and VMP_{S1} , respectively. The initial labor allocation is O_TA units to textiles and O_SA units to steel. The initial wage rate in both industries is w_1 .

Figure 5.15 Effects of a Price Increase



Now suppose the price of steel increases exogenously. The immediate effect will be to raise the value of the marginal product of steel, shifting up VMP_{S1} to VMP_{S2} . The new equilibrium is given at point F. At F, labor allocated to steel production will have risen to $O_S B$, while labor used in textiles will have fallen to $O_T B$. The equilibrium wage increases to w_2 .

The intuition for these changes follows from the underlying dynamic effects. At first, when the price of steel rises, the wage and rental rates remain fixed. This means steel revenue rises while costs remain the same, stimulating an increase in steel profits. Positive profit, in a perfectly competitive market, induces new entry of firms into steel production, expansion of current firms in the industry, or both. To expand, steel must induce workers to move over from textile production. This requires an increase in the wage since labor demand temporarily exceeds labor supply. To prevent all the labor from shifting to steel, the textile industry must raise the wage to its workers as well. As labor shifts from textiles to steel and as the wage rises, the costs of production in steel and textiles rise. In steel, this erodes the temporary profits it was making. Textiles respond to the higher costs by cutting production and releasing workers. Remember, there is no ability to expand capital inputs in steel since we assume steel's capital stock is fixed exogenously in size, and

due to specificity, capital cannot be moved in from the textile industry. In the end, industry profits are driven to zero in both industries once the wage rises sufficiently.

Our prime concern, however, is the effect of the price increase on the factor payments or returns. In other words, how are wages and rental rates on capital affected by the steel price increase? The answer for wages is already shown. We can see that wages rise for workers in both industries. However, we care about not just how the nominal (money) wage changes but, more importantly, how the *real wage* changes. In other words, we need to identify how the purchasing power of wages changes when the price of steel increases. We also want to know how the real rental rates change.

Real Wage Effect

When the price of steel rises from P_{S1} to P_{S2} , the value of the marginal product line shifts up proportionally to the increase in the price. This is because the price of steel enters the value of the marginal product formula multiplicatively—that is,

$VMP_S = P_S MP_S$. The percentage change in the steel price \hat{P}_S is derived in [Figure 5.15 "Effects of a Price Increase"](#) as

$$\hat{P}_S = \frac{DA - EA}{EA} = \frac{DE}{EA}.$$

Here's why. First, the distance DA is the value of the marginal product for labor usage $O_S A$ when the price of steel is P_{S2} . The distance EA is the value of the marginal product for labor usage $O_S A$ when the price of steel is P_{S1} . Thus

$$\frac{DA - EA}{EA} = \frac{P_{S2} MP_S - P_{S1} MP_S}{P_{S1} MP_S} = \frac{P_{S2} - P_{S1}}{P_{S1}} = \hat{P}_S.$$

Note that MP_S cancels out because it is evaluated at the same labor input given by point A .

Similarly, since FB is the equilibrium wage at steel price P_{S2} and CB is the wage at steel price P_{S1} , the percentage change in the equilibrium wage \hat{w} is given by

$$\hat{w} = \frac{FB - CB}{CB} = \frac{FC}{CB}.$$

From [Figure 5.15 "Effects of a Price Increase"](#), it is obvious that $\hat{P}_S > \hat{w}$, which means that the percentage change in the price of steel exceeds the percentage change in the wage rate.

Since in the exercise the price of textiles remains constant, $\hat{P}_T = 0$, we can expand the inequality to

$$\hat{P}_S > \hat{w} > \hat{P}_T.$$

Since $\hat{P}_S > \hat{w}$, this implies that w/P_S , the real wage in terms of steel purchases, decreases. In other words, workers in both industries will be able to buy less steel after the steel price increase than before. However, $\hat{w} > \hat{P}_T$, which implies that w/P_T , the real wage in terms of textile purchases, increases. This means all workers will be able to buy more textiles after the steel price increase than before. In terms of overall well-being, workers will lose in total if they tend to purchase more steel products and fewer textile products. However, if a person's preferences are tilted toward more textiles than steel, then the person may be better off.

Real Rental Effect

When the price of steel rises from P_{S1} to P_{S2} , the rental bill in the steel industry rises from area KEI to area JFH in [Figure 5.15 "Effects of a Price Increase"](#). Since the amount of capital in steel remains fixed, this must mean that the rental rate on steel capital increases. However, simply by looking at the diagram, it is impossible to tell if that increase exceeds or falls short of the percentage change in the price of steel. We'll discuss this issue further.

The rental bill in the textile industry falls from area w_1EG to area w_2FG in [Figure 5.15 "Effects of a Price Increase"](#). Since the amount of capital in steel remains fixed, this must mean that the rental rate on textile capital decreases. Furthermore, since the price of steel increases and the price of textiles stays the same, it must follow that r_T/P_S and r_T/P_T decrease. Therefore, the real rental rate on textile capital must fall with respect to purchases of both goods when the price of steel increases.

Magnification Effect

A definitive ordering of the percentage changes in all goods and factor prices in a two-good SF model was derived mathematically by Jones (1971). See R. W. Jones, "A

Three-Factor Model in Theory, Trade and History," in *Trade, Balance of Payments and Growth*, ed. J. N. Bhagwati, R. W. Jones, R. A. Mundell, and J. Vanek (Amsterdam: North-Holland Publishing Co., 1971). The magnification effect for the SF model is analogous to the magnification effect for prices demonstrated in the Heckscher-Ohlin (H-O) model. It defines an ordering of percentage changes in factor prices induced by changes in the goods' prices. Thus suppose the price of steel rises by a greater percentage than the price of textiles such that $\hat{P}_S > \hat{P}_T$. This may occur if two countries move together in trade or if a trade or domestic policy is changed. Jones showed that the magnification effect in this case would be

$$\hat{r}_S > \hat{P}_S > \hat{w} > \hat{P}_T > \hat{r}_T.$$

Since $\hat{r}_S > \hat{P}_S$ and $\hat{r}_S > \hat{P}_T$, this implies r_S/P_S and r_S/P_T both increase. Thus the real returns to steel capital increase with respect to both goods.

Since $\hat{P}_S > \hat{r}_T$ and $\hat{P}_T > \hat{r}_T$, r_T/P_S and r_T/P_T both decrease. Thus the real returns to textile capital decrease with respect to both goods. Finally, since $\hat{P}_S > \hat{w}$, w/P_S , the real wage in terms of steel purchases, decreases. Thus workers will be able to buy less steel than before. However, $\hat{w} > \hat{P}_T$, which implies that w/P_T , the real wage in terms of textile purchases, increases. This means all workers will be able to buy more textiles than before.

An alternative version of the magnification effect in this model can be written for the case when the price of textiles rises by a greater percentage than the price of steel such that $\hat{P}_T > \hat{P}_S$. The magnification effect in this case becomes

$$\hat{r}_T > \hat{P}_T > \hat{w} > \hat{P}_S > \hat{r}_S.$$

This implies that the real returns to capital in the textile industry increase, and the real returns to capital in the steel industry decrease with respect to purchases of both goods. As before, though, the effect on wages is mixed. Real wages with respect to steel purchases increase, while real wages with respect to textile purchases fall.

Effects of Trade

Since this model is a variation of the H-O model, production technologies are assumed to be identical between countries and trade occurs due to differences in

factor proportions. Since there are ostensibly three factors—labor, textile capital, and steel capital—a country will export those goods that use its relatively abundant factor most intensively. Generally, this model is analyzed by assuming a country conforms to the trade pattern described by the H-O model.

Thus if steel production is capital intensive and the country is capital abundant, then in autarky the price of steel will be relatively lower domestically than abroad, while the price of textiles will be relatively higher. Upon opening trade, the price of steel will begin to rise as steel is exported and the price of textiles will fall as textiles are imported. These price changes are all one needs to apply the magnification effect.

If we assume trade leads to $\hat{P}_S > \hat{P}_T$, then $\hat{r}_S > \hat{P}_S > \hat{w} > \hat{P}_T > \hat{r}_T$. This implies that the return to capital in the export industry (steel) rises, while the return to capital in the import-competing industry (textiles) falls. The return to mobile labor rises with respect to imported goods but falls with respect to export goods.

In contrast, if a country experiences the opposite price change such that $\hat{P}_T > \hat{P}_S$, then the country must be exporting textiles and importing steel. This implies $\hat{r}_T > \hat{P}_T > \hat{w} > \hat{P}_S > \hat{r}_S$. Thus the return to capital in the export industry (textiles) rises, while the return to capital in the import-competing industry (steel) falls. The return to mobile labor rises with respect to imported goods but falls with respect to export goods.

Now we can state more formally and generally that if capital is immobile between industries (or specific to an industry) and if labor is homogeneous and freely mobile between industries, then free trade will cause an increase in the real rents earned by capital in the export industry, a decrease in real rents earned by capital in the import-competing industry, an increase in real wages with respect to purchases of the import goods, and a decrease in real wages with respect to purchases of the export goods.

KEY TAKEAWAYS

- The specific factor (SF) model is a variant of the H-O model that assumes capital is specific to an industry, while labor is freely mobile between industries.
- The Mussa diagram shows how the increase in the price of one product raises wages, raises the rental rate on capital specific to that industry, and lowers the rent on capital specific to the other industry.
- The magnification effect in the SF model demonstrates that the real rent rises in the export industry and falls in the import industry.
- The magnification effect in the SF model demonstrates that real wages in both industries rise with respect to purchases of the import good and fall with respect to purchases of the export good.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe the amount of additional dollars earned from one additional unit of labor input applied in production.
 - b. The value of the marginal product is found by multiplying the marginal product by this variable.
 - c. A condition that is satisfied in the specific factor model at the profit-maximizing level of output.
 - d. The term describing the sum of the total wage bill and the total capital bill.
 - e. Of *increase, decrease, or stay the same*, the effect on the value of the marginal product of grapes when there is a decrease in the market price of grapes in a specific factor model.
 - f. Of *increase, decrease, or stay the same*, the effect on wage bill in the grape industry when there is an increase in the market price of grapes in a specific factor model.
 - g. Of *increase, decrease, or stay the same*, the effect on the equilibrium wage rate when there is a decrease in the market price of one of two goods in a specific factor model.
 - h. The magnification effect for prices in a two-good specific factor model with specific capital and mobile labor when a country opens to trade and exports milk and imports cookies.
 - i. The magnification effect for prices in a two-good specific factor model with specific capital and mobile labor when a country that exports wine and imports cheese moves from free trade to autarky.

5.17 Dynamic Income Redistribution and Trade

LEARNING OBJECTIVE

1. Integrate the results of income redistribution from three separate models: the immobile factor model, the specific factor (SF) model, and the Heckscher-Ohlin (H-O) model.

A number of trade models demonstrate that movements to free trade will cause a redistribution of income. The immobile factor model concludes that income will be redistributed from workers in the import-competing industry to workers in the export industry. The specific factor (SF) model concludes that owners of capital in the export sector will gain at the expense of capital owners in the import-competing sector and that the effects on workers in both industries are ambiguous. The Heckscher-Ohlin (H-O) model demonstrates that income will be redistributed from owners of a country's scarce factor, who will lose, to owners of a country's abundant factor, who will gain.

One of the key distinctions between these models is the degree of factor mobility. The immobile factor model represents one extreme, in which factors are stuck in one industry and cannot move between sectors. The H-O model represents another extreme, in which factors can move freely and costlessly between sectors. The SF model represents an intermediate special case in which one factor is completely immobile and the other is completely mobile.

As was discussed in detail in [Chapter 4 "Factor Mobility and Income Redistribution"](#), [Section 4.2 "Domestic Factor Mobility"](#), different factors of production will likely have different degrees of mobility. Some factors are easily adaptable to other industries. For example, accountants are needed in all businesses, and trucks can be used to transport tomatoes or software. Other factors are so specialized that they cannot be easily adapted for use in other industries. Machinery is often carefully designed for a particular production process and cannot be applied elsewhere.

However, the adaptability of any productive factor is likely to change over time, with mobility rising the longer the amount of time that elapses (see [Chapter 4 "Factor Mobility and Income Redistribution"](#), [Section 4.3 "Time and Factor Mobility"](#)). Thus, if a country were to suddenly liberalize trade, in the very short run—perhaps up to a few weeks—most of the productive factors would not adjust to the change in prices. This is the situation reflected in the immobile factor model.

After a few months or more, the most adaptable factors of production would begin to move from the import-competing sectors to the export sectors, while the least adaptable factors would remain stuck in their respective industries. This situation is characterized by the SF model, in which one factor is freely mobile but the other is immobile. Finally, in the very long run—perhaps after several years or more—we might expect all factors to have adapted to the changed economic conditions, either by moving to another industry or by moving out of productive activity, as with retired workers and capital equipment. This situation is depicted in the H-O model.

Thus, by piecing together the results of these models, we can evaluate how income redistribution is likely to change dynamically over time in response to any shock to the system, such as a movement toward trade liberalization or free trade.

Scenario Setup/Assumptions

Consider a country that produces two goods, which we simply label the import good and the export good, respectively. Production of these two goods requires two factors of production, capital and labor. Assume that the country in question is capital abundant vis-à-vis its trading partner and that the export good is capital intensive relative to the import good. In general, we maintain all the assumptions of the H-O model, with one exception: we will assume that in the short run, capital and labor are completely immobile between industries; in the medium run, labor is freely mobile but capital remains immobile; and in the long run, both labor and capital are freely and costlessly mobile between industries.

We will consider the effects of trade liberalization, although any change that affects the relative prices of the goods can be expected to stimulate similar dynamic effects. Trade liberalization, which in the extreme would be a movement from autarky to free trade, would raise the price of the country's export good and lower the price of its import good. The change in prices sets off the following effects.

Short-Run Effects: Immobile Factor Model

The immobile factor model, beginning in [Chapter 4 "Factor Mobility and Income Redistribution"](#), [Section 4.4 "Immobile Factor Model Overview and Assumptions"](#), was based on a variation of the Ricardian model. As such, the model assumed only one factor of production and different production technologies across countries. The results from that model do carry over into this two-factor, identical technology context, however.

First, consider the transition to the change in output prices. When the price of the export good rises, firms in the export industry will begin to collect more revenue

from sales of their product. Initially, firm profit will begin to rise since the wage rate and rental rate on capital remains fixed. The increase in profit will stimulate the desire to expand production, but production cannot expand by drawing factors from the other industry due to the immobility of factors. Instead, profit-seeking firms within the industry will begin to compete for the capital and labor already in the industry. (Immobility of factors across industries does not mean that factors cannot move between firms within the industry. Recall also that the assumption of perfect competition implies that there are many, many firms operating within an industry.)

Each export firm now has the incentive to lure workers and capital away from other export firms so that it can expand its own production and raise its share of the industry profit. However, the only way to entice factor mobility within the industry is to offer a higher wage and a higher rent. Some factors may now move to other firms, while others may simply negotiate a higher payment from their present employer to induce them to stay. This bidding war will raise both the wage rate and the rental rate to factors employed within the export industry. The bidding war will end once the total factor cost to each firm is equal to revenue and the profit is driven to zero.

In the import industry, firms now face a lower price and hence a lower revenue. Profits will become negative for all firms in the industry. The firms' only options to cut their losses are to contract by laying off workers or to lower the payments to the workers and capital owners. We will assume, for simplicity, that full employment prevails. However, we could easily imagine the bargaining strategy of the firm managers with the workers: "Either we lower your wages or we eliminate your job." Given that factors are assumed to be immobile across industries, there is no hope, at least in the short run, of finding another job. If you are laid off, you could find alternative employment in another firm, but it would only hire you at a lower wage. The assumption of full employment, then, really just means that the price system in the market responds to the excess supply of workers and capital in this industry by lowering factor prices until all the factors are fully employed. Therefore, wages and rents will fall in the import-competing industry until profit in the industry rises to zero and losses are eliminated.

Although it is more difficult to explain intuitively, the real returns to factors in the export industry will rise, while the real returns to factors in the import-competing industry will fall. This means that workers and capital owners in the export industry will have greater purchasing power after trade liberalization, while workers and capitalists in the import-competing industry will be able to buy less.

The final short-run effects are summarized in [Figure 5.16 "Short-Run Factor Income Effects of Trade Liberalization"](#). Both workers and capitalists affiliated with the export industry will benefit from trade liberalization, while workers and capitalists affiliated with the import-competing sector will lose from free trade. Note that income redistribution, at least initially, is based on industry affiliation. What determines who wins and who loses is the industry from which you receive your income.

Figure 5.16 Short-Run Factor Income Effects of Trade Liberalization

	Export Industry	Import Industry
Workers	Gain	Lose
Capitalists	Gain	Lose

Medium-Run Effects: The Specific Factor Model

The SF model is based on a variation of the H-O model. It assumes that one factor, labor, is freely mobile between the two industries, while the second factor, capital, is completely immobile between industries. Although it is unlikely that one factor would move completely before another begins to adjust, the SF model nonetheless is an easily representable intermediate position between the short-run and long-run effects.

First, consider the transition to equilibrium in the SF model. After the final adjustment depicted in the immobile factor model, the wage rate paid to workers in

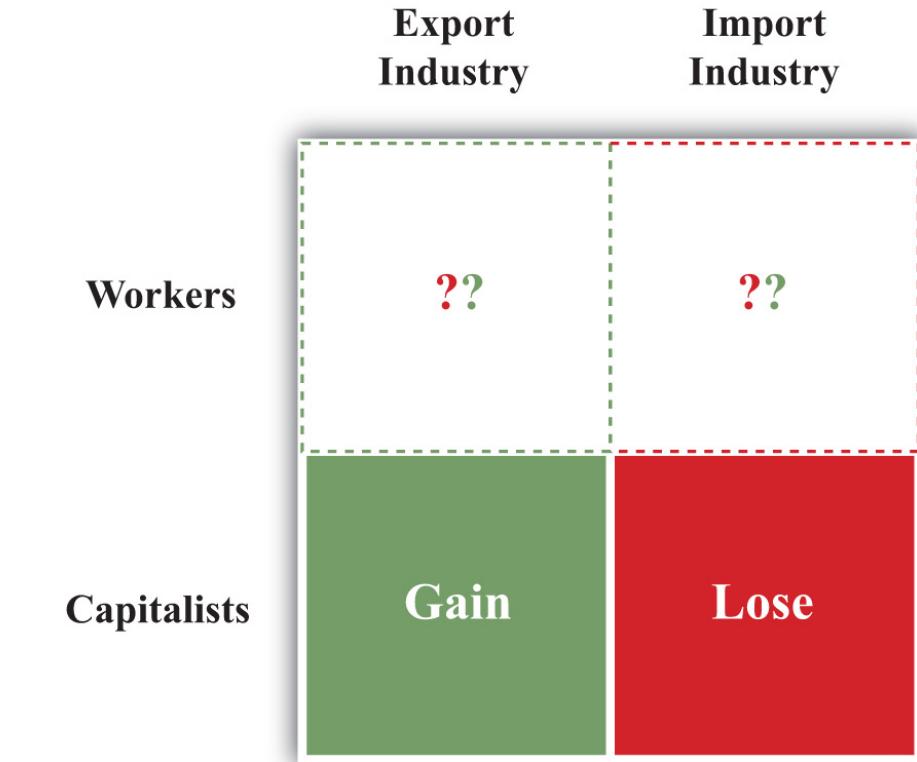
the export industry is higher than the wage paid in the import-competing industry. In the next step of the transition, workers (assumed to be the more readily mobile factor) in the import-competing industry begin to seek ways to obtain a higher wage. This might require additional education or training, or it may require workers to move to another geographical area. In any case, the transition takes time. As workers begin to move across sectors, the supply of labor to the export industry will rise. Profit-seeking firms in that sector will realize that they can temporarily raise their profits by lowering their wage and hiring workers moving in from the other sector. Competition among export firms will eventually lower the wage of all workers in the export industry. Competition within the industry for the specific immobile capital will bid up the rental rate even further than in the short run.

At the same time, the workers fleeing the import-competing sector will reduce the supply of labor there. Import firms will bid among themselves for the remaining workers to maintain output and profit, which will raise the wage paid to workers in this sector. With declining output, the demand for capital will fall, causing an even further drop in the rental rate paid to capital owners.

When the final adjustment of labor across sectors is complete, the wage paid to workers in both industries will be equal. Capital remains in its original sector, but changing prices and outputs affect its sectoral demand. The rental rate paid to capital owners in the export industry will remain higher than that obtained before trade liberalization and will increase relative to the short run. The rental rate for capital owners in the import-competing sector will remain lower than that obtained before trade liberalization.

The magnification effect for prices in this model can be used to assess the real return to factors in the medium-run equilibrium relative to the returns prior to trade liberalization. It shows that the real return to capital owners in the export industry will rise with respect to purchases of both goods, while the real return to capital in the import industry will fall with respect to purchases of both goods. Thus, as shown in [Figure 5.17 "Medium-Run Factor Income Effects of Trade Liberalization"](#), capitalists in the export industry gain and capitalists in the import industry lose.

Figure 5.17 Medium-Run Factor Income Effects of Trade Liberalization



The effect on workers is, in general, ambiguous. The real wage of workers in terms of purchases of the import good rises, while the real wage in terms of the export good falls. For this reason, we place a question mark in [Figure 5.17 "Medium-Run Factor Income Effects of Trade Liberalization"](#) to note the ambiguity. Whether a worker benefits or loses depends, in part, on the worker's preferences. If a worker has a high demand for the import good for which the real wage rises, then the worker may benefit. If, however, a worker has a relatively high demand for the export good, then the worker would lose.

Long-Run Effects: The Heckscher-Ohlin Model

The H-O model assumes that both factors, labor and capital, are freely mobile between the two industries. As such, this corresponds to a long-run outcome after factors fully adjust to the changes in prices.

After the final adjustment depicted in the SF model, the wage rate paid to workers is the same in both industries, but the rental rate on capital in the export industry is higher than the rental rate paid in the import-competing industry. In the next step of the transition, capital owners (assumed to adjust in the final stage) in the

import-competing industry begin to seek ways to obtain higher rents. This might require adapting the capital equipment for use in the export sector or waiting for the capital to fully depreciate and then reinvesting in capital that is usable in the export sector. In any case, the transition takes time. As capital begins to move across sectors, the supply of capital in the export industry will rise. Profit-seeking firms in that sector will realize that they can temporarily raise their profits by lowering their rental and hiring capital moving in from the other sector. Capital owners already in the export sector will have to begin accepting a lower rental payment to avoid being laid off. After all, firm owners can argue that there is no need to keep paying the higher rental rates when there is now a flood of accessible capital streaming in from the import sector.

In the import-competing sector, the loss of capital to the export sector makes capital relatively scarcer in the import sector. This leads to competition among firms for the capital that remains and forces up the price of capital in the import industry. Capital will cease to move between the two industries when the price of capital is equal in both sectors.

As the capital adjusts between industries, outputs and wage rates also adjust. Because the expanding export industry is capital intensive, its demand of capital per worker is greater than the amount of capital per worker that the labor-intensive import industry is able to give up. This implies that the relative demand for capital is higher in the transition to the long-run equilibrium, which results in an increase in the equilibrium rental rate. However, the relative demand for workers in the transition is lower, and this causes a reduction in the equilibrium wage rate. See J. P. Neary, "Short-Run Capital Specificity and the Pure Theory of International Trade," *Economic Journal* 88, no. 351 (1978): 488–510, for an excellent description of the transition between the medium-run effects in the SF model and the long-run effects in the H-O model.

The magnification effect for prices in the H-O model reveals the real returns to the factors relative to those obtained prior to trade liberalization. The effect shows that the equilibrium rental rate rises by a greater percentage than the percentage changes in the two goods' prices, indicating an absolute increase in the real rental rate for all capital owners. The effect also shows that the percentage change in the wage rate is less than the changes in both output prices, indicating an absolute reduction in the purchasing power of all workers' wages. Since capital is the country's relatively abundant factor vis-à-vis the rest of the world and labor is its relatively scarce factor, the general conclusion is that a country's abundant factor gains from trade liberalization, while a country's scarce factor loses. This result is indicated in [Figure 5.18 "Long-Run Factor Income Effects of Trade Liberalization"](#). Note that capital owners are shown to gain regardless of whether their capital is

used in the expanding export sector or the declining import sector. Similarly, all workers lose, even those working in the expanding export sector.

Figure 5.18 Long-Run Factor Income Effects of Trade Liberalization

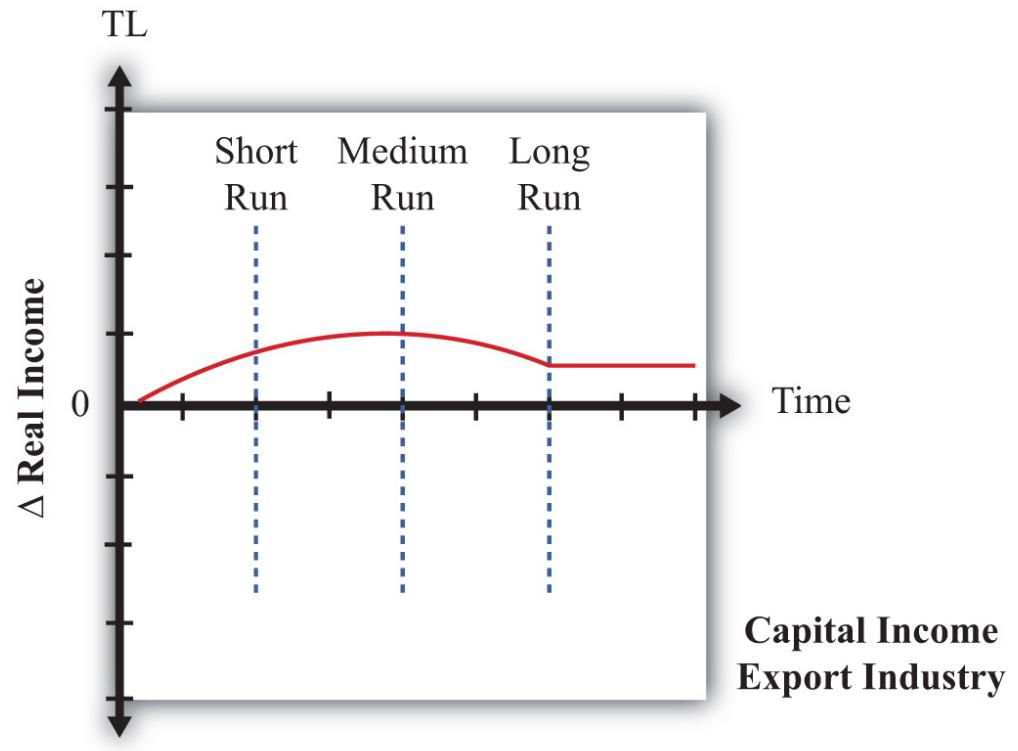
	Export Industry	Import Industry
Workers	Lose	Lose
Capitalists	Gain	Gain

Factor Rewards over Time

Now let's consider the dynamic impact of trade liberalization on factor returns. [Figure 5.19 "Dynamic Export-Capital Income Effects of Trade Liberalization"](#), [Figure 5.20 "Dynamic Import-Capital Income Effects of Trade Liberalization"](#), [Figure 5.21 "Dynamic Export-Labor Income Effects of Trade Liberalization"](#), and [Figure 5.22 "Dynamic Import-Labor Income Effects of Trade Liberalization"](#) depict the changes in real income that might arise over time as a result of trade liberalization. We look at the following four factors in turn: (1) capital owners initially in the export industry, (2) capital owners initially in the import industry, (3) workers originally in the export industry, and (4) workers originally in the import industry. On the horizontal axis in [Figure 5.19 "Dynamic Export-Capital Income Effects of Trade Liberalization"](#), we plot time, with the initial time labeled TL to indicate when trade liberalization occurs. The equilibria that arise in the short run, medium run, and long run are depicted by the vertical blue dotted lines. On the vertical axis, we plot the change in real income, with zero representing the initial preliberalization level.

When the graph is above zero, it indicates an increase in real income; when the graph is below zero, it represents a decrease in real income.

Figure 5.19 Dynamic Export-Capital Income Effects of Trade Liberalization

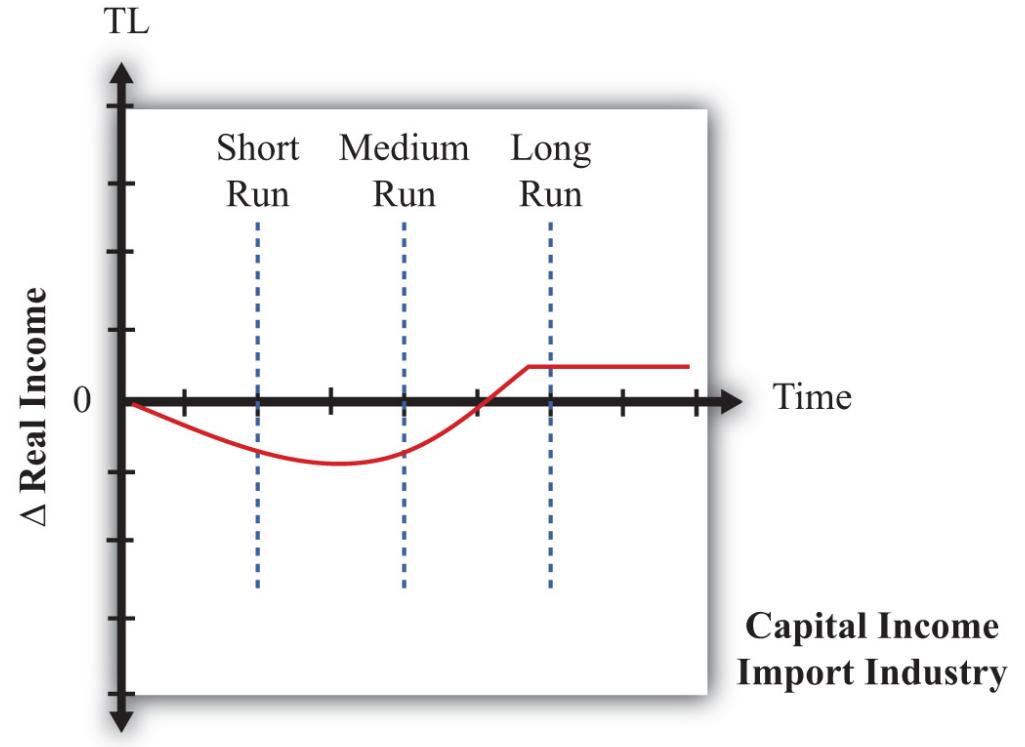


First, consider the owners of capital in the export industry before trade liberalization occurs. The series of models suggests that they will gain in the short run, gain in the medium run, and gain in the long run. However, the transition stories suggest that initial short-run gains would be followed by an increase in these gains in the medium run, but owners would suffer a reduction in their gains in the long term. The dynamic path might look like the red line depicted in [Figure 5.19 "Dynamic Export-Capital Income Effects of Trade Liberalization"](#). Note that although the factor gains throughout the transition, the magnitude of its gains varies.

The models suggest that owners of capital initially in the import industry lose in the short run, will lose further in the medium run, but will ultimately gain in the long run. Its dynamic path might look like the red line in [Figure 5.20 "Dynamic Import-Capital Income Effects of Trade Liberalization"](#). Since this factor experiences both gains and losses, one way to evaluate whether these factor owners are indeed better off would be to calculate the present discounted value of this stream of costs and

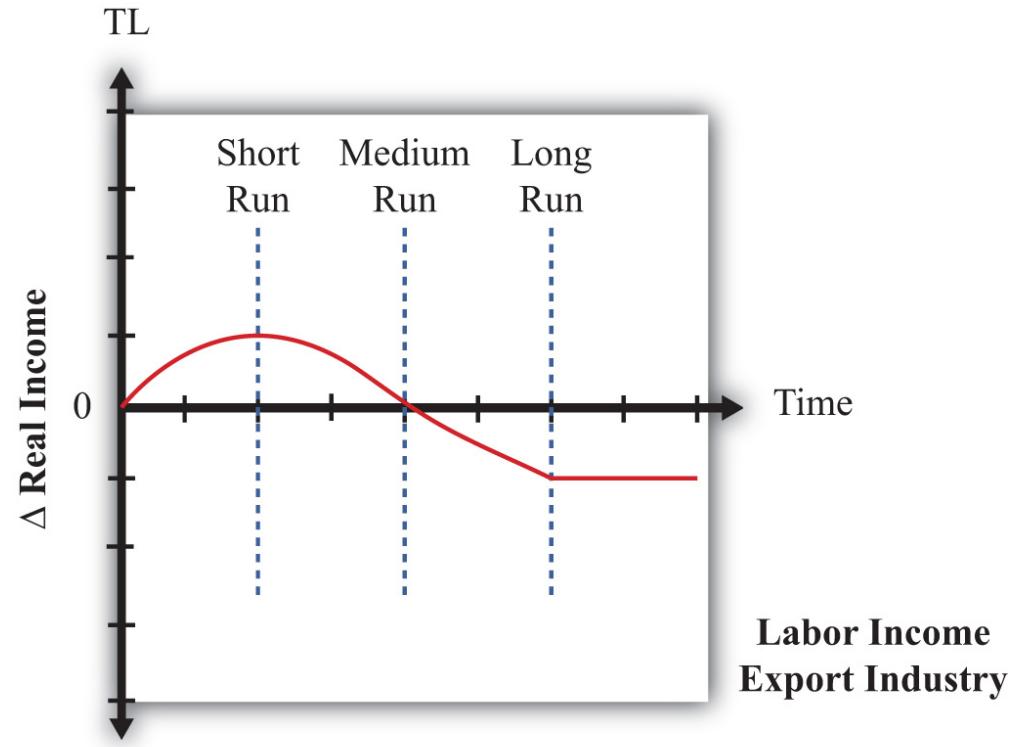
benefits. If the period of losses is sufficiently large or lasts long enough or if the discount rate is high and the person is myopic, the present value may be negative. Otherwise, the discounted value will be positive.

Figure 5.20 Dynamic Import-Capital Income Effects of Trade Liberalization



The models suggest that workers who initially work in the export industry will experience gains in real income in the short run, followed by ambiguous effects in the medium run, followed by losses in the long run. The dynamic path might look like the red line shown in [Figure 5.21 "Dynamic Export-Labor Income Effects of Trade Liberalization"](#). The path is drawn such that the medium-run effect is zero, but the path could be either positive or negative at that point. The present value of this stream of benefits and costs could be positive or negative. If the short-run benefits are sufficiently large or last long enough or if the discount rate is high, then the present value would be positive. Otherwise, the present value is negative.

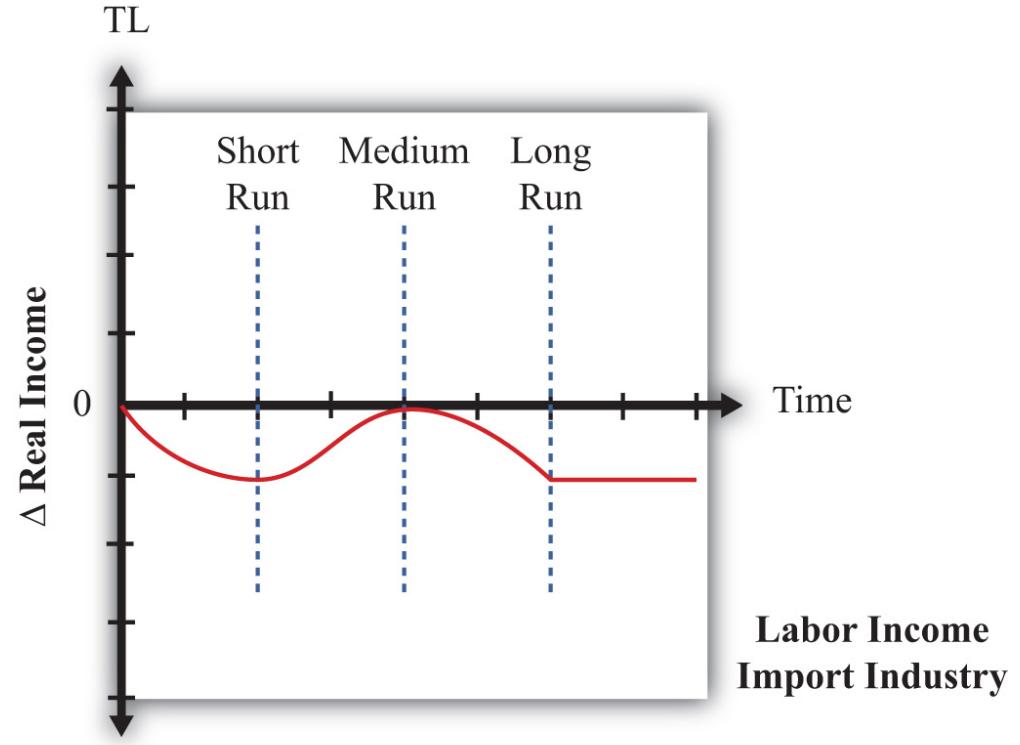
Figure 5.21 Dynamic Export-Labor Income Effects of Trade Liberalization



**Labor Income
Export Industry**

Finally, the models suggest that workers initially in the import sector will lose in the short run, experience ambiguous effects in the medium run, and ultimately lose in the long run. Its dynamic time path may look like the red line in [Figure 5.22 "Dynamic Import-Labor Income Effects of Trade Liberalization"](#). We have set the medium-run effects to zero, but they conceivably could be positive or negative. The present value of this path is likely to be negative even if the factor experiences some medium-run gains.

Figure 5.22 *Dynamic Import-Labor Income Effects of Trade Liberalization*



In summary, the models suggest that the effects of trade liberalization on factor income are rather complex. Some factors will benefit in the short, medium, and long run. Some will lose in all periods. However, some factors will benefit in the short run and lose in the long run, while others will lose in the short run and gain in the long run. The determinants of these paths are whether income is from a relatively abundant factor or from a relatively scarce factor and which industry the factor is employed in before the liberalization occurs.

KEY TAKEAWAYS

- Three models of trade can be interpreted as representing three time frames of factor adjustment to a new equilibrium.
- The immobile factor model represents a very short-run perspective. The specific factor model represents a medium-run perspective. The H-O model represents a long-run perspective.
- By piecing together the results of the models in a dynamic adjustment story, one can demonstrate greater complexity in the effects on factor incomes as time passes after an adjustment to free trade. Most factors will experience changing real income effects as the degree of factor mobility rises over time.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade liberalization on the real income of nationally scarce workers in the import-competing industry in the long run in a dynamic model in which both factors are immobile between industries in the short run, capital is immobile in the medium run, and both factors are mobile in the long run.
 - b. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade liberalization on the real income of nationally abundant capital in the import-competing industry in the short run in a dynamic model in which both factors are immobile between industries in the short run, capital is immobile in the medium run, and both factors are mobile in the long run.
 - c. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade liberalization on the real income of nationally abundant capital in the export industry in the medium run in a dynamic model in which both factors are immobile between industries in the short run, capital is immobile in the medium run, and both factors are mobile in the long run.
 - d. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade liberalization on the real income of nationally scarce capital in the export industry in the long run in a dynamic model in which both factors are immobile between industries in the short run, capital is immobile in the medium run, and both factors are mobile in the long run.
 - e. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade liberalization on the real income of nationally scarce capital in the export industry in the short run in a dynamic model in which both factors are immobile between industries in the short run, capital is immobile in the medium run, and both factors are mobile in the long run.
 - f. Of *increase, decrease, stay the same, or ambiguous*, this is the effect of trade liberalization on the real income of nationally abundant labor in the export industry in the medium run in

a dynamic model in which both factors are immobile between industries in the short run, capital is immobile in the medium run, and both factors are mobile in the long run.

Chapter 6

Economies of Scale and International Trade

One important motivation for international trade is the efficiency improvements that can arise because of the presence of economies of scale in production. Although economists wrote about these effects long ago, models of trade developed after the 1980s introduced economies of scale in creative new ways and became known as the “New Trade Theory.”

In this chapter, the barest essentials of economies of scale models are developed to explain the rationale for trade with this production feature. The chapter also presents the monopolistic competition model of trade that incorporates an obvious feature of the real world—namely, the presence of heterogeneous goods.

6.1 Chapter Overview

LEARNING OBJECTIVE

1. Learn the basic rationale for economies-of-scale models with international trade.

Another major reason that international trade may take place is the existence of economies of scale (also called increasing returns to scale) in production.

Economies of scale¹ means that production at a larger scale (more output) can be achieved at a lower cost (i.e., with economies or savings). When production within an industry has this characteristic, specialization and trade can result in improvements in world productive efficiency and welfare benefits that accrue to all trading countries.

Trade between countries need not depend on country differences under the assumption of economies of scale. Indeed, it is conceivable that countries could be identical in all respects and yet find it advantageous to trade. For this reason, economies-of-scale models are often used to explain trade among countries like the United States, Japan, and the European Union. For the most part, these countries, and other developed countries, have similar technologies, similar endowments, and to some extent similar preferences. Using classical models of trade (e.g., Ricardian, Heckscher-Ohlin), these countries would have little reason to engage in trade. Yet trade between the developed countries makes up a significant share of world trade. Economies of scale can provide an answer for this type of trade.

Another feature of international trade that remains unexplained with classical models is the phenomenon of intraindustry trade. A quick look at the aggregate trade data reveals that many countries export and import similar products. For example, the United States imports and exports automobiles, imports and exports machine tools, imports and exports steel, and so on. To some extent, intraindustry trade arises because many different types of products are aggregated into one category. For example, many different types of steel are produced, from flat-rolled to specialty steels. It may be that production of some types of steel requires certain resources or technologies in which one country has a comparative advantage.

Another country may have the comparative advantage in another type of steel. However, since all these types are generally aggregated into one export or import category, it could appear as if the countries are exporting and importing “identical”

1. The feature of many production processes in which the per-unit cost of producing a product falls as the scale of production rises.

products when in actuality they are exporting one type of steel and importing another type.

Nevertheless, it is possible to explain intraindustry trade in a model that includes economies of scale and differentiated products even when there are no differences in resources or technologies across countries. This model is called the monopolistic competition model. Its focus is on consumer demand for a variety of characteristics embodied in the goods sold in a product category. In this model, advantageous trade in differentiated products can occur even when countries are very similar in their productive capacities.

KEY TAKEAWAYS

- The presence of economies of scale in production represents another reason countries may trade with each other.
- Economies-of-scale models are used to explain intraindustry trade—that is, trade between countries with similar characteristics, like the United States and Canada.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”

- a. The term used to describe when both exports and imports of a good occur in the same industry.
- b. The term used to describe production in which the unit cost falls as the size of the industry becomes larger.
- c. Models incorporating this assumption about production are used to explain trade between countries with similar characteristics.

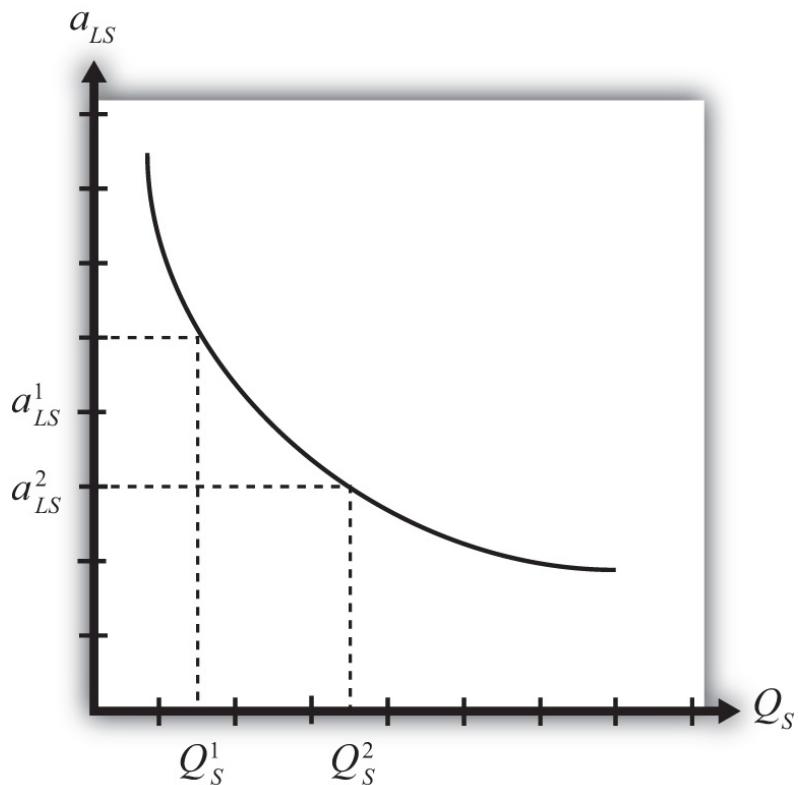
6.2 Economies of Scale and Returns to Scale

LEARNING OBJECTIVE

- Distinguish economies of scale from increasing returns to scale.

Economies of scale in production means that production at a larger scale (more output) can be achieved at a lower cost (i.e., with economies or savings). A simple way to formalize this is to assume that the unit labor requirement in the production of a good is a function of the level of output produced. In [Figure 6.1 "Unit-Labor Requirement with Economies of Scale"](#), we present a graph of the unit labor requirement in steel production as a function of the scale (level of output) of production. At production level Q_S^1 , the unit labor requirement is given by a_{LS}^1 . If production were to rise to Q_S^2 , then the unit labor requirement would fall to a_{LS}^2 . This means that at the higher level of output, it requires less labor (i.e., fewer resources or a lower cost) per unit of output than it required at the smaller scale.

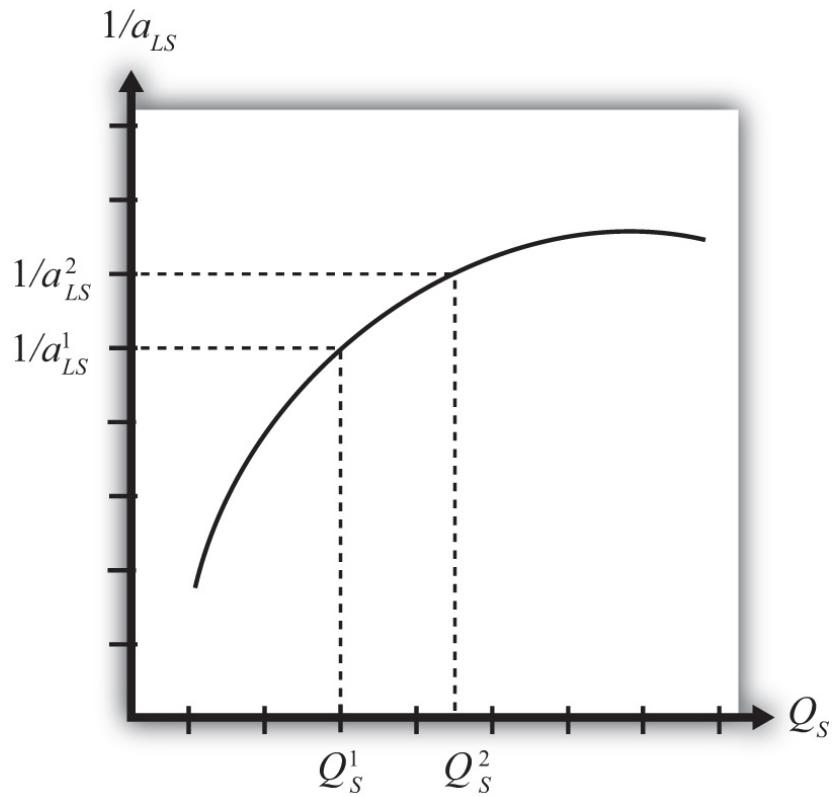
Figure 6.1 Unit-Labor Requirement with Economies of Scale



A secondary assumption is that the additional savings (or economies) fall as the scale increases. Graphically, this means that the slope of the curve in [Figure 6.1 "Unit-Labor Requirement with Economies of Scale"](#) becomes less negative as the scale of production (output) rises. Economists sometimes refer to this feature by saying the function is concave to the origin; that is, it is bowed inward. The reason this assumption is made is because it seems to correspond to what is observed in the world. We expect that the degree of cost savings will be largest in the earliest stages of production, when labor division is likely to be the easiest and most effective. This assumption, although a realistic feature, is not necessary to explain trade, however.

With a simple adjustment, it is possible to show that **increasing returns to scale**² in production means that an increase in resource usage by, say, x percent results in an increase in output by more than x percent. In [Figure 6.2 "Productivity with Increasing Returns to Scale"](#), we plot labor productivity in steel production when production exhibits increasing returns to scale. This curve is derived by plotting the reciprocal of the unit labor requirement (i.e., $1/a_{LS}$) for each output level in [Figure 6.2 "Productivity with Increasing Returns to Scale"](#).

Figure 6.2 Productivity with Increasing Returns to Scale



2. The feature of many production processes in which the productivity of a product increases as the scale of production rises.

Note that as output (scale) increases from Q_S^1 to Q_S^2 , labor productivity (given by the reciprocal of the unit labor requirement) also rises. In other words, output per unit of labor input increases as the scale of production rises, hence increasing returns to scale.

Another way to characterize economies of scale is with a decreasing average cost curve. Average costs, AC , are calculated as the total costs to produce output Q , $TC(Q)$, divided by total output. Thus $AC(Q) = TC(Q)/Q$. When average costs decline as output increases, it means that it becomes cheaper to produce the average unit as the scale of production rises, hence resulting in economies of scale.

Economies of scale are most likely to be found in industries with large fixed costs in production. Fixed costs are those costs that must be incurred even if production were to drop to zero. For example, fixed costs arise when large amounts of capital equipment must be put into place even if only one unit is to be produced and if the costs of this equipment must still be paid even with zero output. In this case, the larger the output, the more the costs of this equipment can be spread out among more units of the good. Large fixed costs and hence economies of scale are prevalent in highly capital-intensive industries such as chemicals, petroleum, steel, automobiles, and so on.

Economies of Scale and Perfect Competition

It is worth noting that the assumption of economies of scale in production can represent a deviation from the assumption of perfectly competitive markets. In most perfectly competitive models, it is assumed that production takes place with constant returns to scale (i.e., no economies). This means that the unit cost of production remains constant as the scale of production increases. When that assumption is changed, it can open up the possibility of positive profits and strategic behavior among firms. Because there are numerous ways to conceive of strategic interactions between firms, there are also numerous models and results that could be obtained. To avoid some of these problems, a number of models have been developed that retain some of the key features of perfect competition while allowing for the presence of economies of scale as well.

KEY TAKEAWAYS

- Economies of scale refers to the feature of many production processes in which the per-unit cost of producing a product falls as the scale of production rises.
- Increasing returns to scale refers to the feature of many production processes in which productivity per unit of labor rises as the scale of production rises.
- The introduction of economies of scale in production in a model is a deviation from perfect competition when positive economic profits are allowed to prevail.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe rising productivity in an industry as the scale of production increases.
 - b. The assumption about scale economies normally made in perfect competition.
 - c. The term used to describe total production costs per unit of output.
 - d. The assumption made about scale economies if a 10 percent increase in factor inputs causes a 10 percent increase in output.
 - e. The assumption made about scale economies if a 10 percent increase in factor inputs causes a 20 percent increase in output.

6.3 Gains from Trade with Economies of Scale: A Simple Explanation

LEARNING OBJECTIVE

1. Learn how a simple model can show the gains from trade when production involves economies of scale.

The main reason the presence of economies of scale can generate trade gains is because the reallocation of resources can raise world productive efficiency. To see how, we present a simple example using a model similar to the Ricardian model.

Basic Assumptions

Suppose there are two countries, the United States and France, producing two goods, clothing and steel, using one factor of production, labor. Assume the production technology is identical in both countries and can be described with the production functions in [Table 6.1 "Production of Clothing"](#).

Table 6.1 Production of Clothing

United States	France
$Q_C = \frac{L_C [\text{hrs}]}{a_{LC} \left[\frac{\text{hrs}}{\text{rack}} \right]}$	$Q_C^* = \frac{L_C^*}{a_{LC}}$

where

Q_C = quantity of clothing produced in the United States

L_C = amount of labor applied to clothing production in the United States

a_{LC} = unit labor requirement in clothing production in the United States and France (hours of labor necessary to produce one rack of clothing)

United States	France
<p>*All starred variables are defined in the same way but refer to the production process in France.</p>	

Note that since production technology is assumed to be the same in both countries, we use the same unit labor requirement in the U.S. and French production functions.

Production of steel. The production of steel is assumed to exhibit economies of scale in production (see [Table 6.2 "Production of Steel"](#)).

Table 6.2 Production of Steel

United States	France
$Q_S = \frac{L_S [\text{hrs}]}{a_{LS}(Q_S) \left[\frac{\text{hrs}}{\text{tons}} \right]}$	$Q_S^* = \frac{L_S^*}{a_{LS}(Q_S^*)}$
<p>where</p>	
<p>Q_S = quantity of steel produced in the United States</p>	
<p>L_S = amount of labor applied to steel production in the United States</p>	
<p>$a_{LS}(Q_S)$ = unit labor requirement in steel production in the United States (hours of labor necessary to produce one ton of steel)</p>	
<p>*All starred variables are defined in the same way but refer to the production process in France.</p>	

Note that it is assumed that the unit labor requirement is a function of the level of steel output in the domestic industry. More specifically, we will assume that the unit labor requirement falls as industry output rises.

Resource constraint. The production decision is how to allocate labor between the two industries. We assume that labor is homogeneous and freely mobile between industries. The labor constraints are given in [Table 6.3 "Labor Constraints"](#).

Table 6.3 Labor Constraints

United States	France
$L_C + L_W = L$	$L_C^* + L_W^* = L^*$

where
 L = labor endowment

When the resource constraint holds with equality, it implies that the resource is fully employed.

Demand. We will assume that the United States and France have identical demands for the two products.

A Numerical Example

We proceed much as David Ricardo did in presenting the argument of the gains from specialization in one's comparative advantage good. First, we will construct an autarky equilibrium in this model assuming that the two countries are identical in every respect. Then we will show how an improvement in world productive efficiency can arise if one of the two countries produces all the steel that is demanded in the world.

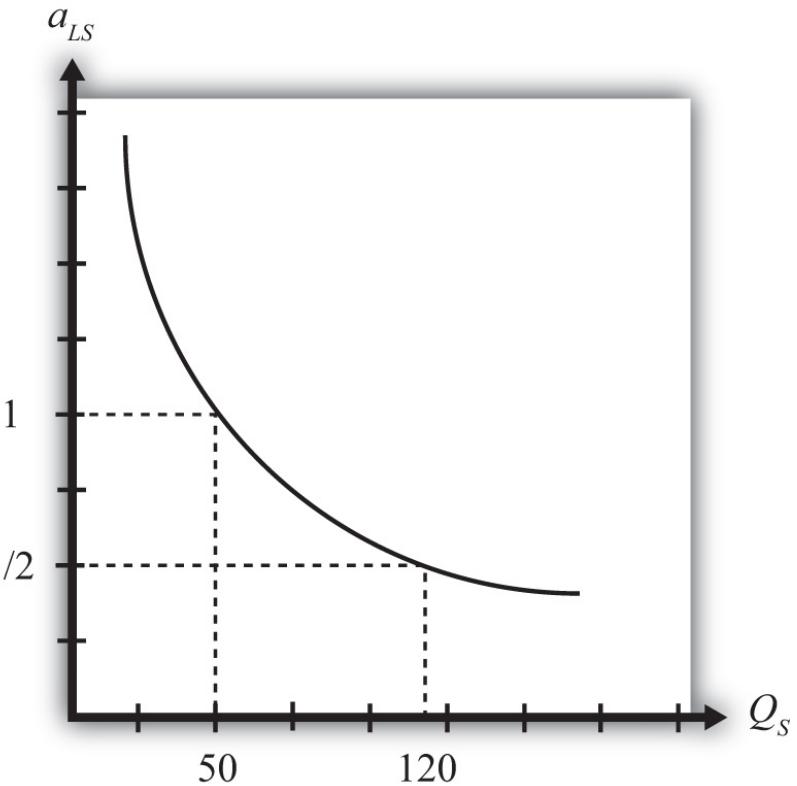
Suppose the exogenous variables in the two countries take the values in [Table 6.4 "Initial Exogenous Variable Values"](#).

Table 6.4 Initial Exogenous Variable Values

United States	$a_{LC} = 1$	$L = 100$
France	$a_{LC}^* = 1$	$L^* = 100$

Let the unit labor requirement for steel vary as shown in [Figure 6.3 "Economies of Scale: Numerical Example"](#). The graph shows that when fifty tons of steel are produced by the economy, the unit labor requirement is one hour of labor per ton of steel. However, when 120 tons of steel are produced, the unit labor requirement falls to half an hour of labor per ton of steel.

Figure 6.3 Economies of Scale: Numerical Example



An Autarky Equilibrium

The United States and France, assumed to be identical in all respects, will share identical autarky equilibria. Suppose the equilibria are such that production of steel in each country is fifty tons. Since at fifty tons of output, the unit labor requirement is one, it means that the total amount of labor used in steel production is fifty hours. That leaves fifty hours of labor to be allocated to the production of clothing. The production of clothing has a unit labor requirement of one also, meaning that the total output of clothing is fifty racks. The autarky production and consumption levels are summarized in [Table 6.5 "Autarky Production/Consumption"](#).

Table 6.5 Autarky Production/Consumption

	Clothing (Racks)	Steel (Tons)
United States	50	50
France	50	50
World Total	100	100

The problem with these initial autarky equilibria is that because demands and supplies are identical in the two countries, the prices of the goods would also be identical. With identical prices, there would be no incentive to trade if trade suddenly became free between the two countries.

Gains from Specialization

Despite the lack of incentive to trade in the original autarky equilibria, we can show, nevertheless, that trade could be advantageous for both countries. All that is necessary is for one of the two countries to produce its good with economies of scale and let the other country specialize in the other good.

For example, suppose we let France produce 120 tons of steel. This is greater than the 100 tons of world output of steel in the autarky equilibria. Since the unit labor requirement of steel is one-half when 120 tons of steel are produced by one country, the total labor can be found by plugging these numbers into the production function. That is, since $Q_S^* = L_S^*/a_{LS}^*$, $Q_S^* = 120$ and $a_{LS}^* = \frac{1}{2}$, it must be that $L_S^* = 60$. In autarky, it took 100 hours of labor for two countries to produce 100 tons of steel. Now it would take France 60 hours to produce 120 tons. That means more output with less labor.

If France allocates its remaining forty hours of labor to clothing production and if the United States specializes in clothing production, then production levels in each country and world totals after the reallocation of labor would be as shown in Table 6.6 "Reallocated Production".

Table 6.6 Reallocated Production

	Clothing (Racks)	Steel (Tons)
United States	100	0
France	40	120

	Clothing (Racks)	Steel (Tons)
World Total	140	120

The important result here is that it is possible to find a reallocation of labor across industries and countries such that world output of both goods rises. Or in other words, there is an increase in world productive efficiency.

If output of both goods rises, then surely it must be possible to find terms of trade such that both countries would gain from trade. For example, if France were to export sixty tons of steel and import thirty racks of clothing, then each country would consume seventy units of clothing (twenty more than in autarky) and sixty tons of steel (ten more than in autarky).

The final conclusion of this numerical example is that when there are economies of scale in production, then free trade, after an appropriate reallocation of labor, can improve national welfare for both countries relative to autarky. The welfare improvement arises because concentrating production in the economies-of-scale industry in one country allows one to take advantage of the productive efficiency improvements.

Some Noteworthy Features

Some features of the economies-of-scale model make it very different from the other models of trade, such as the Ricardian or Heckscher-Ohlin models. For example, it is possible to show that countries that are identical in every respect might nevertheless find it advantageous to trade. Thus it is not always differences between countries that stimulate trade. In this case, it is a feature of the production process (i.e., economies of scale) that makes trade gains possible.

Second, this economies-of-scale model cannot predict which country would export which good. It doesn't matter which country produces all the economies-of-scale good. As long as one country does so and trades it with the rest of the world, trade gains are possible. Also, it may not matter whether your country ends up producing the economies-of-scale good or not because both countries will realize the benefits as long as an appropriate terms of trade arises.

Despite these differences with other models, the main similarity is that gains from trade arise because of an improvement in productive efficiency. By reallocating resources between industries within countries, it is possible to produce more output with the same amount of resources. This remains the prime motivation in support of free trade.

KEY TAKEAWAYS

- By shifting production in one country to production of the good that exhibits economies of scale and shifting production toward the other good in the other country, it is possible to raise total output in the world with the same total resources.
- Countries that are identical in every respect can benefit from trade in the presence of economies of scale.
- Countries that are identical would have no natural incentive to trade because there would be no price differences between countries.
- A simple economies-of-scale model does not predict which country would export which good.

EXERCISE

1. Suppose there are two countries with the same production technologies. Let labor productivity in butter production be ten pounds per hour at all levels of output and productivity in gun production be one-half of a gun per hour when gun production is less than ten and two-thirds of a gun per hour when production is ten or more. Suppose each country has fifty hours of labor and in autarky produces eight guns.
 - a. Calculate how many pounds of butter each country produces in autarky.
 - b. What is the total world output of guns and butter in autarky?

Next, suppose Country A produces all the guns in the world while Country B specializes in butter production.

- c. Calculate the quantity of butter produced by Country A and Country B.
- d. What is total world output of guns and butter now?
- e. Identify a terms of trade (guns for butter) that will assure that each country is at least as well off after trade as before.

6.4 Monopolistic Competition

LEARNING OBJECTIVE

1. Identify the basic features of a monopolistic competition model.

Monopolistic competition³ refers to a market structure that is a cross between the two extremes of perfect competition and monopoly. The model allows for the presence of increasing returns to scale in production and for differentiated (rather than homogeneous or identical) products. However, the model retains many features of perfect competition, such as the presence of many, many firms in the industry and the idea that free entry and exit of firms in response to profit would eliminate economic profit among the firms. As a result, the model offers a somewhat more realistic depiction of many common economic markets. The model best describes markets in which numerous firms supply products that are each slightly different from that supplied by its competitors. Examples include automobiles, toothpaste, furnaces, restaurant meals, motion pictures, romance novels, wine, beer, cheese, shaving cream, and much more.

The model is especially useful in explaining the motivation for **intraindustry trade**⁴—that is, trade between countries that occurs within an industry rather than across industries. In other words, the model can explain why some countries export and import automobiles simultaneously. This type of trade, although frequently measured, is not readily explained in the context of the Ricardian or Heckscher-Ohlin models of trade. In those models, a country might export wine and import cheese, but it would never export and import wine at the same time.

The model demonstrates not only that intraindustry trade may arise but also that national welfare can be improved as a result of international trade. One reason for the improvement in welfare is that individual firms produce larger quantities, which, because of economies of scale in production, leads to a reduction in unit production costs. This means there is an improvement in productive efficiency. The second reason welfare improves is that consumers are able to choose from a greater variety of available products with trade as opposed to autarky.

3. A market structure that is a cross between the two extremes of perfect competition and monopoly.
4. Trade between countries that occurs within the same industry; for example, when a country exports and imports automobiles.

KEY TAKEAWAYS

- A monopolistic competition market represents a cross between a monopoly market and a perfectly competitive market.
- Intraindustry trade refers to trade within a particular industry. An example is a country that both exports and imports cars.
- A monopolistic competition model can explain why intraindustry trade may occur between countries.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The type of market structure that mixes assumptions from perfect competition with assumptions from monopoly models.
 - b. The term used to describe two-way trade in identical or similar products.
 - c. The term used to describe nonhomogeneous goods produced by different firms within the same industry.

6.5 Model Assumptions: Monopolistic Competition

LEARNING OBJECTIVE

1. Compare the assumptions of a monopolistic competition model with monopoly and perfect competition assumptions.

A monopolistically competitive market has features that represent a cross between a perfectly competitive market and a monopolistic market (hence the name). The following are some of the main assumptions of the model:

1. Many, many firms produce in a monopolistically competitive industry. This assumption is similar to that found in a model of perfect competition.
2. Each firm produces a product that is differentiated (i.e., different in character) from all other products produced by the other firms in the industry. Thus one firm might produce a red toothpaste with a spearmint taste, and another might produce a white toothpaste with a wintergreen taste. This assumption is similar to a monopoly market that produces a unique (or highly differentiated) product.
3. The differentiated products are imperfectly substitutable in consumption. This means that if the price of one good were to rise, some consumers would switch their purchases to another product within the industry. From the perspective of a firm in the industry, it would face a downward-sloping demand curve for its product, but the position of the demand curve would depend on the characteristics and prices of the other substitutable products produced by other firms. This assumption is intermediate between the perfectly competitive assumption in which goods are perfectly substitutable and the assumption in a monopoly market in which no substitution is possible.

Consumer demand for differentiated products is sometimes described using two distinct approaches: the love-of-variety approach and the ideal variety approach. The love-of-variety approach assumes that each consumer has a demand for multiple varieties of a product over time. A good example of this would be restaurant meals. Most consumers who eat out frequently will also switch between restaurants, one day eating at a Chinese restaurant, another day at a Mexican restaurant, and so on. If all consumers share the same love of variety, then the aggregate market will sustain demand for many

varieties of goods simultaneously. If a utility function is specified that incorporates a love of variety, then the well-being of any consumer is greater the larger the number of varieties of goods available. Thus the consumers would prefer to have twenty varieties to choose from rather than ten.

The ideal variety approach assumes that each product consists of a collection of different characteristics. For example, each automobile has a different color, interior and exterior design, engine features, and so on. Each consumer is assumed to have different preferences over these characteristics. Since the final product consists of a composite of these characteristics, the consumer chooses a product closest to his or her ideal variety subject to the price of the good. In the aggregate, as long as consumers have different ideal varieties, the market will sustain multiple firms selling similar products. Therefore, depending on the type of consumer demand for the market, one can describe the monopolistic competition model as having consumers with heterogeneous demand (ideal variety) or homogeneous demand (love of variety).

4. There is free entry and exit of firms in response to profits in the industry. Thus firms making positive economic profits act as a signal to others to open up similar firms producing similar products. If firms are losing money (making negative economic profits), then, one by one, firms will drop out of the industry. Entry or exit affects the aggregate supply of the product in the market and forces economic profit to zero for each firm in the industry in the long run. (Note that the long run is defined as the period of time necessary to drive the economic profit to zero.) This assumption is identical to the free entry and exit assumption in a perfectly competitive market.
5. There are economies of scale in production (internal to the firm). This is incorporated as a downward-sloping average cost curve. If average costs fall when firm output increases, it means that the per-unit cost falls with an increase in the scale of production. Since monopoly markets can arise when there are large fixed costs in production and since fixed costs result in declining average costs, the assumption of economies of scale is similar to a monopoly market.

These main assumptions of the monopolistically competitive market show that the market is intermediate between a purely competitive market and a purely monopolistic market. The analysis of trade proceeds using a standard depiction of equilibrium in a monopoly market. However, the results are reinterpreted in light of these assumptions. Also, it is worth mentioning that this model is a partial

equilibrium model since there is only one industry described and there is no interaction across markets based on an aggregate resource constraint.

KEY TAKEAWAYS

- The monopolistic competition assumptions of many firms, free entry and exit, and imperfect substitutability between products are most similar to a perfectly competitive market.
- The monopolistic competition assumptions of differentiated products, economies of scale, and imperfect substitutability between products are most similar to a monopoly market.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The demand assumption in which each consumer has a demand for multiple varieties of a product over time.
 - b. The demand assumption in which each consumer has a demand for different sets of characteristics of a particular product type.
 - c. This is a standard perfect competition assumption indicating what new firms do in response to positive profit in an industry.
 - d. This is a standard perfect competition assumption indicating what existing firms do in response to negative profit in an industry.
 - e. The production feature that is present when a firm’s average cost curve is downward sloping.
 - f. Of *many* or *few*, this is the assumption made about the number of firms in a monopolistically competitive industry.
 - g. The long-run value of firm profit in a monopolistically competitive industry.

6.6 The Effects of Trade in a Monopolistically Competitive Industry

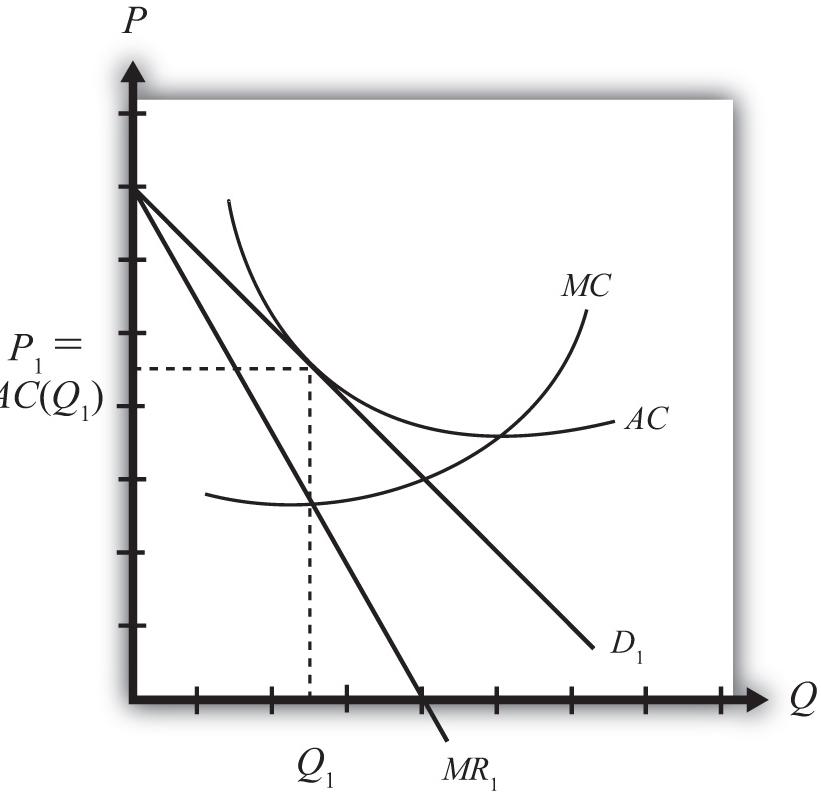
LEARNING OBJECTIVES

1. Use a monopoly diagram for a representative monopolistically competitive firm to depict a long-run equilibrium.
2. Understand how the market equilibrium changes upon opening to free trade.

Assume that there are two countries, each with a monopolistically competitive industry producing a differentiated product. Suppose initially that the two countries are in autarky. For convenience, we will assume that the firms in the industry are symmetric relative to the other firms in the industry. Symmetry implies that each firm has the same average and marginal cost functions and that the demand curves for every firm's product are identical, although we still imagine that each firm produces a product that is differentiated from all others. (Note that the assumptions about symmetry are made merely for tractability. It is much simpler to conceive of the model results when we assume that all firms are the same in their essential characteristics. However, it seems likely that these results would still be obtained even if firms were asymmetric.)

In [Figure 6.4 "Firm Equilibrium in Monopolistic Competition"](#), we depict a market equilibrium for a representative firm in the domestic industry. The firm faces a downward-sloping demand curve (D_1) for its product and maximizes profit by choosing that quantity of output such that marginal revenue (MR_1) is equal to marginal cost (MC). This occurs at output level Q_1 for the representative firm. The firm chooses the price for its product, P_1 , that will clear the market. Notice that the average cost curve (AC) is just tangent to the demand curve at output Q_1 . This means that the unit cost at Q_1 is equal to the price per unit—that is, $P_1 = AC(Q_1)$, which implies that profit is zero. Thus the firm is in a long-run equilibrium since entry or exit has driven profits to zero.

Figure 6.4 Firm Equilibrium in Monopolistic Competition



Keep in mind that this is the equilibrium for just one of many similar firms producing in the industry. Also imagine that the foreign market (which is also closed to trade) has a collection of firms that are also in a long-run equilibrium initially.

Next, suppose whatever barriers to trade that had previously existed are suddenly and immediately removed—that is, suppose the countries move from autarky to free trade. The changes that ultimately arise will be initiated by the behavior of consumers in the market. Recall that market demand can be described using a love-of-variety approach or an ideal variety approach.

In the love-of-variety approach, the removal of trade barriers will increase the number of varieties consumers have to choose from. Since consumer welfare rises as the number of varieties increases, domestic consumers will shift some of their demand toward foreign varieties, while foreign consumers will shift some of their demand toward domestic varieties.

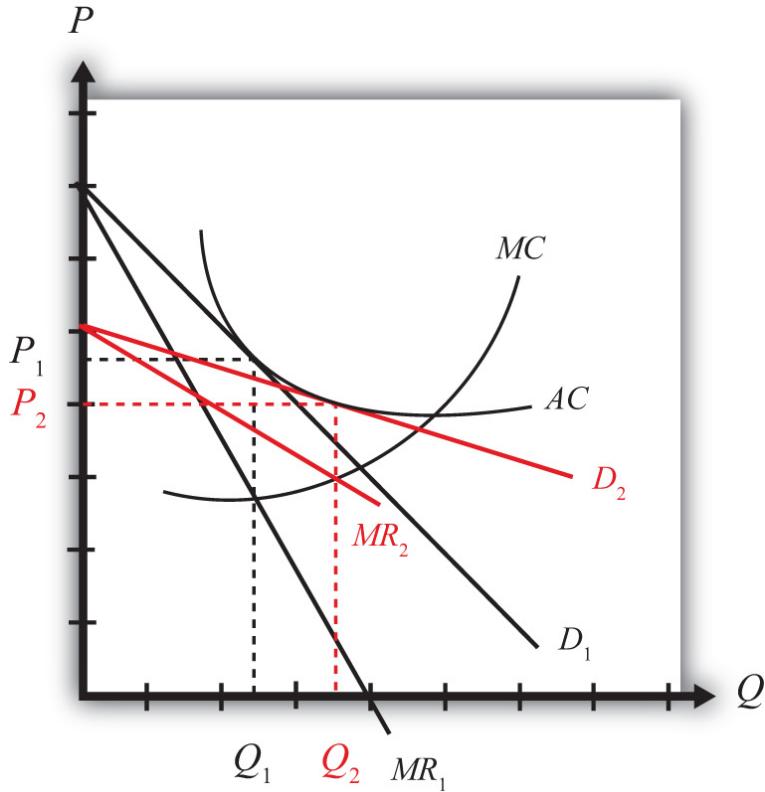
In the ideal variety approach, some domestic consumers will likely discover a more ideal variety produced by a foreign firm. Similarly, some foreign consumers will find a more ideal variety produced by a domestic firm.

In either case, domestic demand by domestic consumers will fall, while domestic demand by foreign consumers will rise. Similarly, foreign demand by foreign consumers will fall, while foreign demand by domestic consumers will rise. Note that this is true even if all the prices of all the goods in both countries are initially identical. In terms of [Figure 6.4 "Firm Equilibrium in Monopolistic Competition"](#), trade will cause the demand curve of a representative firm to shift out because of the increase in foreign demand but will cause the demand curve to shift back in because of the reduction in domestic demand. Since these two effects push the demand curve in opposite directions, the final effect will depend on the relative sizes of these effects.

Regardless of the size of these effects, the removal of trade barriers would cause intraindustry trade to arise. Each country would become an exporter and an importer of differentiated products that would be classified in the same industry. Thus the country would export and import automobiles, toothpaste, clothing, and so on. The main cause of this result is the assumption that consumers, in the aggregate at least, have a demand for variety.

However, two effects can be used to isolate the final equilibrium after trade is opened. First, the increase in the number of varieties available to consumers implies that each firm's demand curve will become more elastic (or flatter). The reason is that consumers become more price sensitive. Since there are more varieties to choose from, a \$1 increase in price of one variety will now lead more consumers to switch to an alternative brand (since there are more close substitutes available), and this will result in a larger decrease in demand for the original product. Second, free entry and exit of firms in response to profits will lead to a zero-profit equilibrium for all remaining firms in the industry.

Figure 6.5 Firm Equilibrium Before and After Trade



The final equilibrium for the representative firm is shown in [Figure 6.5 "Firm Equilibrium Before and After Trade"](#). Keep in mind that these same effects are occurring for every other firm in the industry, both domestically and in the foreign country. The demand curve shifts from D_1 to D_2 and the marginal revenue from MR_1 to MR_2 as a result of trade. The firm's cost curves remain the same. Entry or exit of firms causes the final demand curve to be tangent to the firm's average cost curve, but since the demand curve is more elastic (or flatter), the tangency occurs down and to the right of the autarky intersection. In the end, firm output rises from Q_1 to Q_2 and the price charged in the market falls from P_1 to P_2 . Although individual firm output rises for each firm, we cannot tell in this model setup whether industry output has risen. In the adjustment to the long-run zero-profit equilibrium, entry (or more likely exit) of firms would occur. If some firms exit, then it remains uncertain whether fewer firms, each producing more output, would raise or lower industry output.

KEY TAKEAWAYS

- A market equilibrium for a representative firm in a monopolistically competitive (MC) market displays an output level such that $MR = MC$ and establishes a price such that $P = AC$.
- When trade opens up between two countries that have MC markets, the consumer demand for variety inspires trade.
- Trade in an MC market increases the total number of varieties available to each consumer and causes market demand for each product to become more elastic.
- The free trade equilibrium in an MC market results in a higher quantity produced for each firm and a lower market price than before trade.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The type of cost derived by dividing total cost by total output.
 - b. The type of market demand (elastic or inelastic) that would arise if demand were very responsive to changes in the price.
 - c. This is the relationship between the demand curve and the average cost curve in equilibrium in a monopolistically competitive market.
 - d. The position along the average cost curve where the marginal cost curve intersects in a monopolistically competitive market.
 - e. This is the relationship between the market price and the average cost in equilibrium in a monopolistically competitive market.
 - f. The profit-maximizing condition in a monopolistically competitive market.
 - g. Of *increase, decrease, or stay the same*, this is the effect of international trade on the output of a representative firm in a monopolistically competitive industry.
 - h. Of *increase, decrease, or stay the same*, this is the effect of international trade on the output price of a representative firm in a monopolistically competitive industry.

6.7 The Costs and Benefits of Free Trade under Monopolistic Competition

LEARNING OBJECTIVES

1. Identify the reasons why consumers gain from trade in a monopolistically competitive market.
2. Understand that the movement to free trade in a monopolistically competitive market may cause losses to some individuals under more realistic assumptions.

The Benefits of Free Trade

Welfare of individual consumers who purchase the representative product will be enhanced for three main reasons. First, trade increases the number of varieties of products for consumers to choose from. Second, free trade reduces the price of every variety sold in the market. Third, free trade may increase the supply of products in other markets and result in lower prices for those products.

1. If the product is such that an individual consumer seeks to purchase a product closest to her ideal variety, then presumably with more varieties available, more consumers will be able to purchase more products closer to their ideal. For these consumers welfare will be improved. Other consumers, however, may not be affected by the increase in varieties. If, for example, the new varieties that become available are all more distant from one's ideal than the product purchased in autarky, then one would continue to purchase the same product in free trade. In this case, the increase in variety does not benefit the consumer.

If the product is one in which consumers purchase many different varieties over time (love of variety), then because trade will increase the number of varieties available to each consumer, trade will improve every consumer's welfare. Of course, this is based on the assumption that every consumer prefers more varieties to less. Thus regardless of whether the product is characterized by the ideal variety or the love-of-variety approach, free trade, by increasing the number of varieties, will increase aggregate consumer welfare.

2. The second effect of trade for consumers is that the price of all varieties of the product will fall. The prices fall because trade allows a firm to produce further down its average cost curve, which means that

it lowers its per-unit cost of production. This implies that each product is being produced more efficiently. Competition in the industry, in turn, forces profit to zero for each firm, which implies that the efficiency improvements are passed along to consumers in the form of lower prices.

3. Finally, the improvement in productive efficiency for each firm may lead to a reduction in the use of resources in the industry. This effect would occur if industry output falls or if output does not rise too much. Although the use of resources per unit produced falls, total output by each firm rises. Thus it is uncertain whether an individual firm would have to lay off workers and capital or whether it would need to hire more. However, even if it hired more, the possibility that some firms would drop out of business in the adjustment to the long-run equilibrium might mean that industry resource usage falls. If resource usage does fall and capital and labor are laid off, then in a general equilibrium system (which has not been explicitly modeled here), these resources would be moved into other industries. Production in those industries would rise, leading to a reduction in the prices of those products. Thus free trade in the monopolistically competitive industry can lead to a reduction in prices of completely unrelated industries.

The Costs of Free Trade

There are two potential costs of free trade in this model. The first involves the potential costs of adjustment in the industry. The second involves the possibility that more varieties will increase transaction costs. Each cost requires modification of the basic assumptions of the model in a way that conforms more closely with the real world. However, since these assumption changes are not formally included in the model, the results are subject to interpretation.

1. The movement to free trade requires adjustment in the industry in both countries. Although firm output rises, productive efficiency rises as well. Thus it is possible that each firm will need to lay off resources—labor and capital—in moving to free trade. Even if each firm did not reduce resources, it is possible (indeed likely) that some firms will be pushed out of business in moving to the long-run free trade equilibrium. It is impossible to identify which country's firms would close; however, it is likely to be those firms that lose more domestic customers than they gain in foreign customers or firms that are unable or unwilling to adjust the characteristics of their product to serve the international market rather than the domestic market alone. For firms that close, all the capital and labor employed will likely suffer through an adjustment process. The costs would involve the opportunity cost of

lost production, unemployment compensation costs, search costs associated with finding new jobs, emotional costs of being unemployed, costs of moving, and so on. Eventually, these resources are likely to be reemployed in other industries. The standard model assumption is that this transition occurs immediately and without costs. In reality, however, the adjustment process is likely to be harmful to some groups of individuals.

2. A second potential cost of free trade arises if one questions the assumption that more variety is always preferred by consumers. Consider for a moment a product in which consumers seek their ideal variety. A standard (implicit) assumption in this model is that consumers have perfect information about the prices and characteristics of the products they consider buying. In reality, however, consumers must spend time and money to learn about the products available in a market. For example, when a consumer considers the purchase of an automobile, part of the process involves a search for information. One might visit dealerships and test-drive selected cars, purchase magazines that offer evaluations, or talk to friends about their experiences with different automobiles. All these activities involve expending resources—time and money—and thus represent what we could call a “transactions cost” to the consumer.

Before we argued that because trade increases the number of varieties available to each consumer, each consumer is more likely to find a product that is closer to his or her ideal variety. In this way, more varieties may increase aggregate welfare. However, the increase in the number of varieties also increases the cost of searching for one's ideal variety. More time will now be needed to make a careful evaluation. One could reduce these transaction costs by choosing to evaluate only a sample of the available products. However, in this case, a psychological cost might also arise because of the inherent uncertainty about whether the best possible choice was indeed made. Thus, welfare would be diminished among consumers to the extent that there are increased transaction costs because of the increase in the number of varieties to evaluate.

The Net Welfare Effects of Trade

The welfare effects under the basic assumptions of the model are entirely positive. Improvements in productive efficiency arise as firms produce further down along their average cost curves in free trade. Consumption efficiency is raised because consumers are able to buy the products at lower prices and have a greater variety to choose from.

Potential costs arise in the model only if we introduce the additional assumptions of adjustment costs or transactions costs. The net welfare effect in the presence of adjustment and transactions costs will still be positive if the production and consumption efficiency effects are larger.

KEY TAKEAWAYS

- Consumers benefit from trade in a monopolistically competitive (MC) market because they can consume a greater variety of goods at a lower price.
- Free trade in an MC market may also lower the prices of products in other markets if reduced resource usage results in a shift to other industries causing an increase in supply and thereby a lower price.
- Because some firms may close when an MC market moves to free trade, some of those resources may suffer costs of adjustment.
- Consumer transaction costs to identify the most ideal variety may rise with an increase in the number of varieties available in free trade.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, this is the effect of international trade on the number of varieties of a good available to consumers in a monopolistically competitive market.
 - b. Of *increase, decrease, or stay the same*, this is the effect of international trade on the price of a good available to consumers in a monopolistically competitive market.
 - c. Of *increase, decrease, or stay the same*, this is the effect of international trade on productive efficiency of firms remaining in business in a monopolistically competitive market.
 - d. The two costs associated with adjustment to a trading equilibrium in a monopolistically competitive market.
 - e. Of *positive, negative, or the same*, this is the net welfare effect of international trade in a monopolistically competitive market under the standard assumptions.

Chapter 7

Trade Policy Effects with Perfectly Competitive Markets

Governments have long intervened in international trade by collecting taxes, or tariffs, on imported goods. Tariffs have a long history since they are one of the easiest ways for governments to collect revenue. However, tariffs have a number of other effects besides generating government revenue; they also affect the success of business and the well-being of consumers. And because tariffs affect the volume of trade between countries, they also affect businesses and consumers abroad.

This chapter examines, in detail, the effects of a tariff. However, it also examines the impacts of the many other types of trade policies that governments have applied historically, including import quotas, export quotas, export taxes, and export subsidies.

The effects are considered under one set of standard assumptions—namely, in the case when markets are perfectly competitive.

7.1 Basic Assumptions of the Partial Equilibrium Model

LEARNING OBJECTIVE

1. Identify the basic assumptions of a simple partial equilibrium trade model.

This section analyzes the price and welfare effects of trade policies using a partial equilibrium model under the assumption that markets are perfectly competitive.

1. Assume there are two countries, the United States and Mexico. The analysis can be generalized by assuming one of the countries is the rest of the world.
2. Each country has producers and consumers of a tradable good, wheat. The analysis can be generalized by considering broad classes of products, like manufactured goods, or services.
3. Wheat is a homogeneous good. All wheat from Mexico and the United States is perfectly substitutable in consumption.
4. The markets are perfectly competitive.
5. We assume that the two countries are initially trading freely. One country implements a trade policy and there is no response or retaliation by the other country.

The Meaning of Partial Equilibrium

In **partial equilibrium**¹ analysis, the effects of policy actions are examined only in the markets that are directly affected. Supply and demand curves are used to depict the price effects of policies. Producer and consumer surplus is used to measure the welfare effects on participants in the market. A partial equilibrium analysis either ignores effects on other industries in the economy or assumes that the sector in question is very, very small and therefore has little if any impact on other sectors of the economy.

1. An economic analysis in which the effects are examined only in the markets that are directly affected. Supply and demand curves for the market of interest are typically used in a partial equilibrium analysis.

In contrast, a general equilibrium analysis incorporates the interaction of import and export sectors and then considers the effects of policies on multiple sectors in the economy. It uses offer curves to depict equilibria and measures welfare with aggregate welfare functions or trade indifference curves.

The Large versus Small Country Assumption

Two cases are considered regarding the size of the policy-setting country in international markets. The effects of policies vary significantly depending on the size of a country in international markets.

If the country is a “**large country**²” in international markets, then the country’s imports or exports are a significant share in the world market for the product. Whenever a country is large in an international market, domestic trade policies can affect the world price of the good. This occurs if the domestic trade policy affects supply or demand on the world market sufficiently to change the world price of the product.

If the country is a “**small country**³” in international markets, then the policy-setting country has a very small share in the world market for the product—so small that domestic policies are unable to affect the world price of the good. The small country assumption is analogous to the assumption of perfect competition in a domestic goods market. Domestic firms and consumers must take international prices as given because they are too small for their actions to affect the price.

KEY TAKEAWAYS

- Partial equilibrium analysis uses supply and demand curves in a particular market and ignores effects that occur beyond these markets.
- Large countries are those whose trade volume is significant enough such that large changes in trade flows can affect the world price of the good.
- Small countries are those whose trade volume is *not* significant enough such that any changes in its trade flows will not affect the world price of the good.

2. A country is large if any change in its trade volume for a product is sufficiently large to affect the price of that product in the rest of the world.
3. A country is small if any change in its trade volume for a product is too small to have any effect on the price of that product in the rest of the world.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a country in which domestic policy changes can influence prices in international markets.
 - b. The term used to describe a country in which domestic policy changes cannot influence prices in international markets.
 - c. The term used to describe the substitutability of a good that is homogeneous.
 - d. This type of economic analysis focuses on policy effects within a single market and does not address effects external to the market.

7.2 Depicting a Free Trade Equilibrium: Large and Small Country Cases

LEARNING OBJECTIVES

1. Use supply and demand to derive import demand curves and export supply curves.
2. Combine import demand and export supply curves to depict a free trade equilibrium under the assumption that the countries are large.
3. Use an import demand and export supply diagram to depict a free trade equilibrium under the assumption that the import country is small.

Figure 7.1 "U.S. Wheat Market: Autarky Equilibrium" depicts the supply and demand for wheat in the U.S. market. The supply curve represents the quantity of wheat that U.S. producers would be willing to supply at every potential price for wheat in the U.S. market. The demand curve represents demand by U.S. consumers at every potential price for wheat in the U.S. market. The intersection of demand and supply corresponds to the equilibrium autarky price and quantity in the United States. The price, P_{Aut}^{US} , is the only price that will balance domestic supply with domestic demand for wheat.

Figure 7.1 U.S. Wheat Market: Autarky Equilibrium

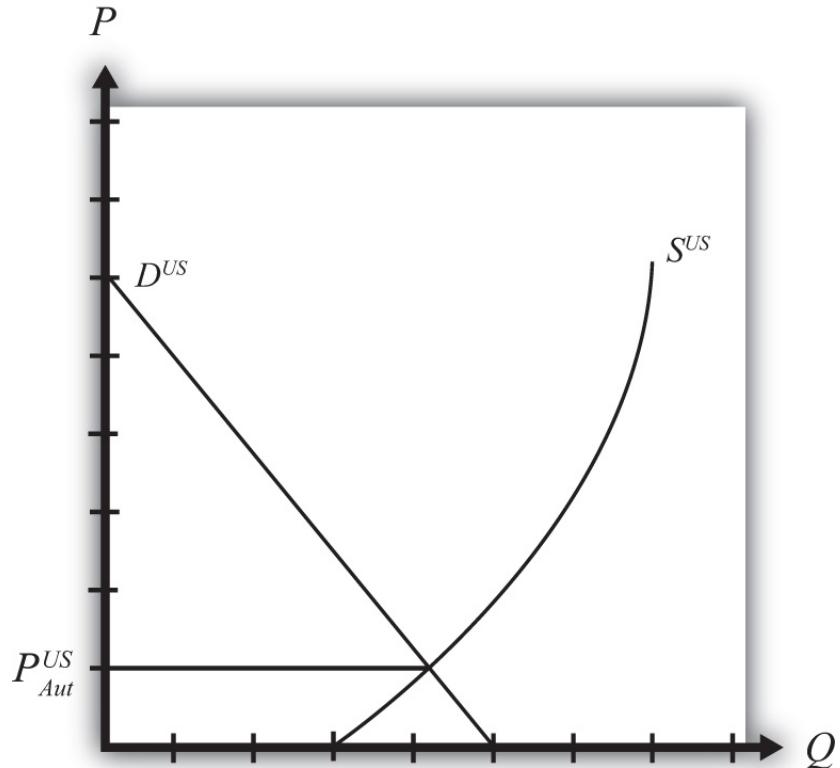
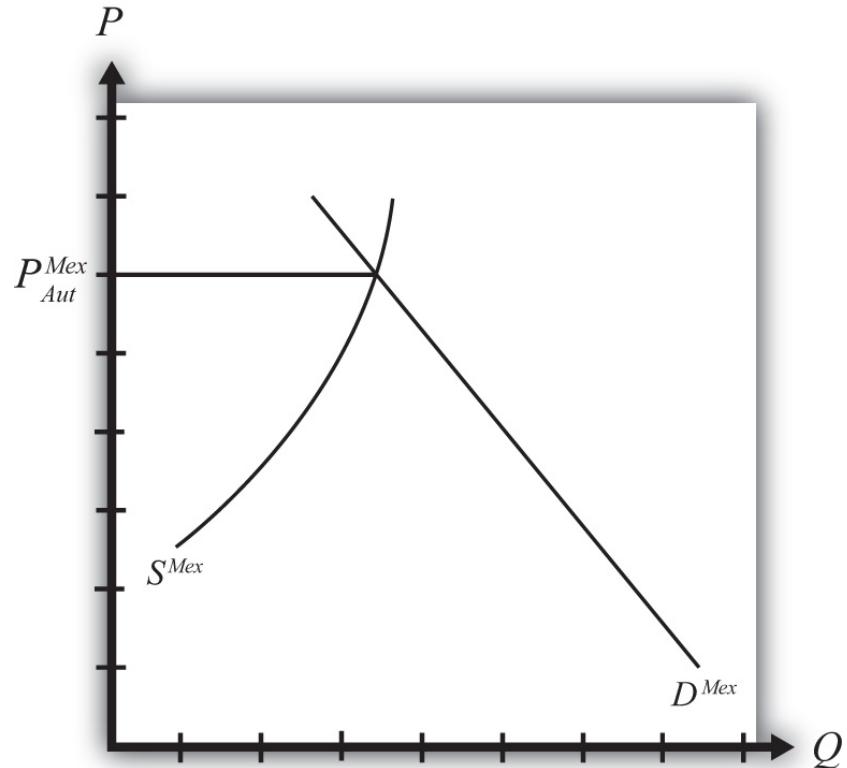


Figure 7.2 "Mexican Wheat Market: Autarky Equilibrium" shows the supply and demand for wheat in the Mexican market. The supply curve represents the quantity of wheat that Mexican producers would be willing to supply at every potential price in the Mexican market. The demand curve represents demand by Mexican consumers at every potential price for wheat in the Mexican market. The intersection of demand and supply corresponds to the equilibrium autarky price and quantity in Mexico. The price, P_{Aut}^{Mex} , is the only price that will balance Mexican supply with demand for wheat.

Figure 7.2 Mexican Wheat Market: Autarky Equilibrium



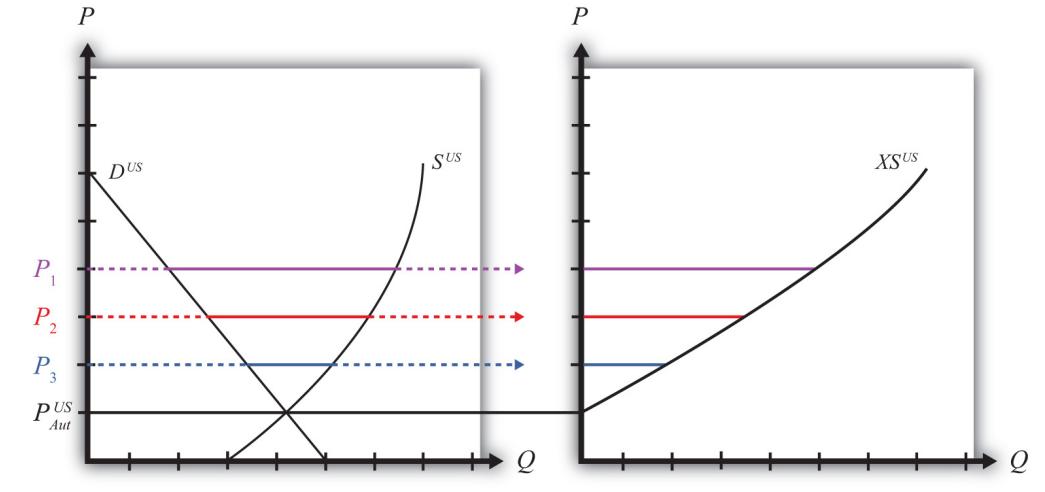
The curves are drawn such that the U.S. autarky price is lower than the Mexican autarky price. This implies that if these two countries were to move from autarky to free trade, the United States would export wheat to Mexico. Once trade is opened, the higher Mexican price will induce profit-seeking U.S. firms to sell their wheat in Mexico, where it commands a higher price initially. As wheat flows into Mexico, the total supply of wheat rises, which will cause the price to fall. In the U.S. market, wheat supply falls because of U.S. exports. The reduced supply raises the equilibrium price in the United States. These prices move together as U.S. exports rise until the prices are equalized between the two markets. The free trade price of wheat, P_{FT} , is shared by both countries.

To derive the free trade price and the quantity traded, we can construct an export supply curve for the United States and an import demand curve for Mexico. Notice that at prices above the autarky price in the United States, there is excess supply of wheat—that is, supply exceeds demand. If we consider prices either at or above the autarky price, we can derive an export supply curve for the United States. The equation for export supply is given by

$$XS^{US}(P^{US}) = S^{US}(P^{US}) - D^{US}(P^{US}),$$

where $XS^{US}(.)$ is the export supply function, $S^{US}(.)$ is the supply function for wheat in the United States, and $D^{US}(.)$ is the demand function for wheat in the United States. Each function is dependent on the U.S. price of wheat, P^{US} .

Figure 7.3 Deriving the U.S. Export Supply Curve



Graphically, **export supply**⁴ is the horizontal difference between the supply and demand curve at every price at and above the autarky price, as shown in [Figure 7.3 "Deriving the U.S. Export Supply Curve"](#). At the autarky price, P_{Aut}^{US} , export supply is zero. At prices P_1 , P_2 , and P_3 , export supply is given by the length of the like-colored line segment. To plot the export supply curve XS^{US} , we transfer each line segment to a separate graph and connect the points, as shown on the right in [Figure 7.3 "Deriving the U.S. Export Supply Curve"](#). The export supply curve gives the quantities the United States would be willing to export if it faced prices above its autarky price.

4. The quantity of a product a country would wish to export at a particular price. The export supply curve is the schedule of export supply at every potential price (usually prices above the country's autarky price).

5. The quantity of a product a country would wish to import at a particular price. The import demand curve is the schedule of import demand at every potential price (usually prices below the country's autarky price).

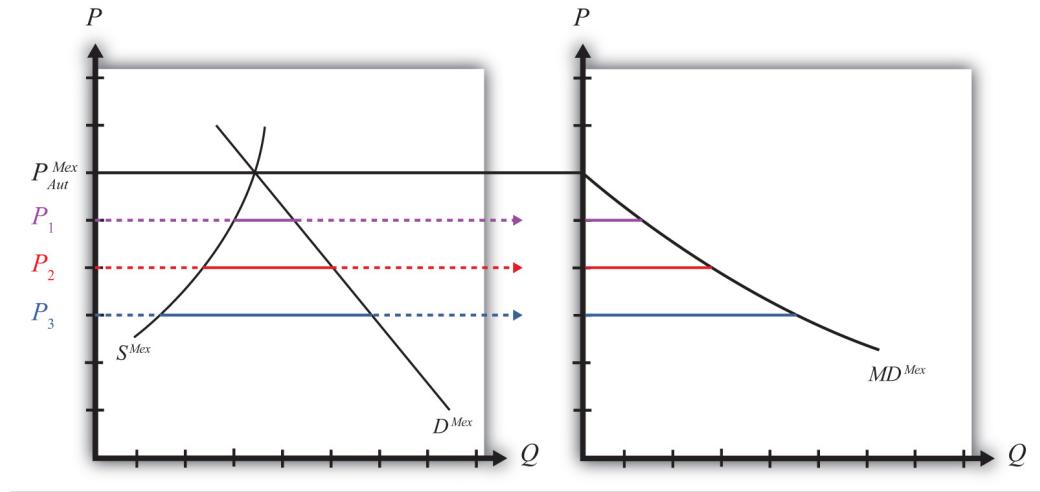
In Mexico, at prices below its autarky price there is excess demand for wheat since demand exceeds supply. If we consider prices either at or below the autarky price, we can derive an import demand curve for Mexico. The equation for import demand is given by

$$MD^{Mex}(P^{Mex}) = D^{Mex}(P^{Mex}) - S^{Mex}(P^{Mex}),$$

where $MD^{Mex}(.)$ is the import demand function, $D^{Mex}(.)$ is the demand function for wheat in Mexico, and $S^{Mex}(.)$ is the supply function for wheat in Mexico. Each function is dependent on the Mexican price of wheat, P^{Mex} . Graphically, **import demand**⁵ is the horizontal difference between the demand and supply curve at every price at and below the autarky price, as shown in [Figure 7.4 "Deriving the](#)

Mexican Import Demand Curve". At the autarky price, P_{Aut}^{Mex} , import demand is zero. At prices P_1 , P_2 , and P_3 , import demand is given by the length of the like-colored line segment. To plot the import demand curve MD^{Mex} , we transfer each line segment to a separate graph and connect the points, as shown on the right in Figure 7.4 "Deriving the Mexican Import Demand Curve". The import demand curve gives the quantities Mexico would be willing to import if it faced prices below its autarky price.

Figure 7.4 *Deriving the Mexican Import Demand Curve*

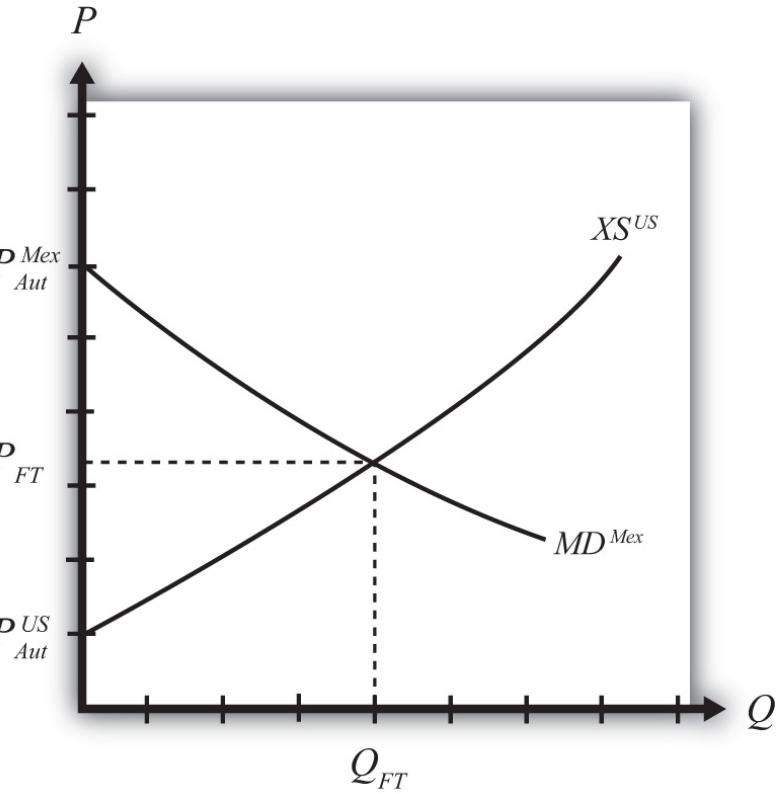


Free Trade Equilibrium: Large Country Case

The intersection of the U.S. export supply with Mexican import demand determines the equilibrium free trade price, P_{FT} , and the quantity traded, Q_{FT} , where $Q_{FT} = XS^{US}(P_{FT}) = MD^{Mex}(P_{FT})$. See Figure 7.5 "Depicting a Free Trade Equilibrium". The free trade price, P_{FT} , must be the price that equalizes the U.S. export supply with Mexican import demand. Algebraically, the free trade price is the price that solves

$$XS^{US}(P_{FT}) = MD^{Mex}(P_{FT})$$

Figure 7.5 Depicting a Free Trade Equilibrium



This implies also that world supply is equal to world demand since

$$S^{US}(P_{FT}) - D^{US}(P_{FT}) = D^{Mex}(P_{FT}) - S^{Mex}(P_{FT})$$

and

$$S^{US}(P_{FT}) + S^{Mex}(P_{FT}) = D^{US}(P_{FT}) + D^{Mex}(P_{FT}).$$

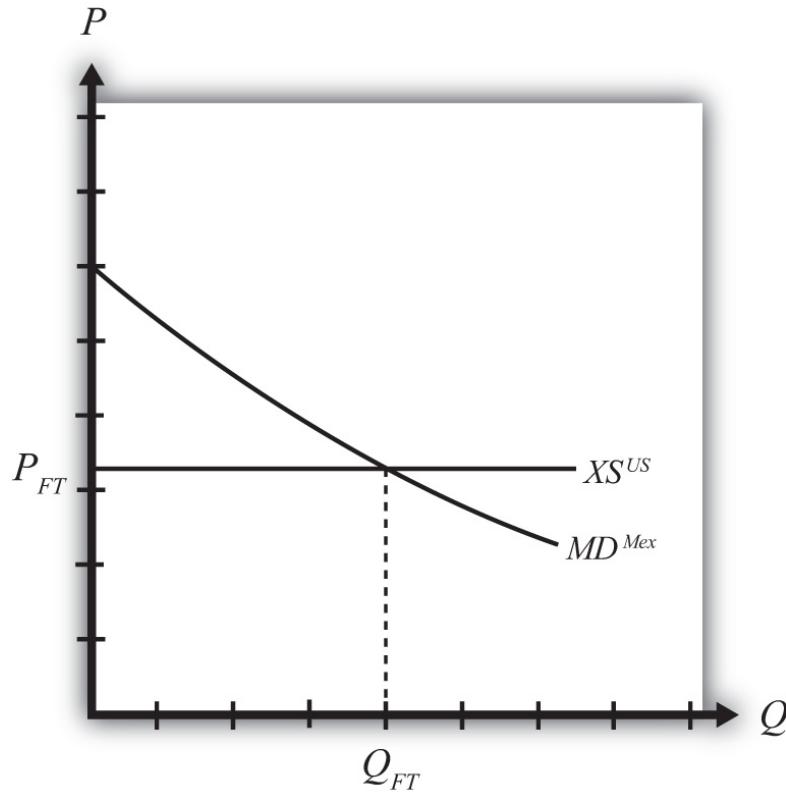
Free Trade Equilibrium: Small Country Case

The small country assumption means that the country's imports are a very small share of the world market—so small that even a complete elimination of imports would have an imperceptible effect on world demand for the product and thus would not affect the world price.

To depict a free trade equilibrium using an export supply and import demand diagram, we must redraw the export supply curve in light of the small country assumption. The assumption implies that the export supply curve is horizontal at

the level of the world price. In this case, we call the importing country small. From the perspective of the small importing country, it takes the world price as exogenous since it can have no effect on it. From the exporter's perspective, it is willing to supply as much of the product as the importer wants at the given world price.

Figure 7.6 Free Trade Equilibrium: Small Country Case



The free trade price, P_{FT} , is the price that prevails in the export, or world, market. The quantity imported into the small country is found as the intersection between the downward-sloping import demand curve and the horizontal export supply curve.

KEY TAKEAWAYS

- Import demand is the excess demand that a country would wish to import from another country if the market price were below the price that equalizes its own supply and demand (i.e., its autarky price).
- Export supply is the excess supply that a country would wish to export to another country if the market price were above the price that equalizes its own supply and demand (i.e., its autarky price).
- When there are only two countries, the free trade price is the one that equalizes one country's import demand with the other's export supply.
- When export supply is equal to import demand, world supply of the product is equal to world demand at the shared free trade price.
- A large importing country faces a downward-sloping export supply curve.
- A small importing country is one that faces a perfectly elastic export supply function.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The price that equalizes one country’s import demand with the other’s export supply.
 - b. Of *higher than, lower than, or equal* to the autarky price in a market, this is the range of prices that would generate positive import demand.
 - c. Of *higher than, lower than, or equal* to the autarky price in a market, this is the range of prices that would generate positive export supply.
 - d. The value of imports of wine in free trade in Country A if Country A’s autarky wine price is equal to the autarky wine price in the rest of the world.
 - e. The term used to describe the horizontal distance between supply and demand at each price below a market autarky price.
 - f. The term used to describe the horizontal distance between supply and demand at each price above a market autarky price.
 - g. The shape of the export supply function faced by a small importing country.

7.3 The Welfare Effects of Trade Policies: Partial Equilibrium

LEARNING OBJECTIVE

1. Measure welfare magnitudes accruing to producers and consumers in a partial equilibrium model.

A partial equilibrium analysis distinguishes between the welfare of consumers who purchase a product and the producers who produce it. Consumer welfare is measured using consumer surplus, while producer welfare is measured using producer surplus. Revenue collected by the government is assumed to be redistributed to others. Government revenue is either spent on public goods or is redistributed to someone in the economy, thus raising someone's welfare.

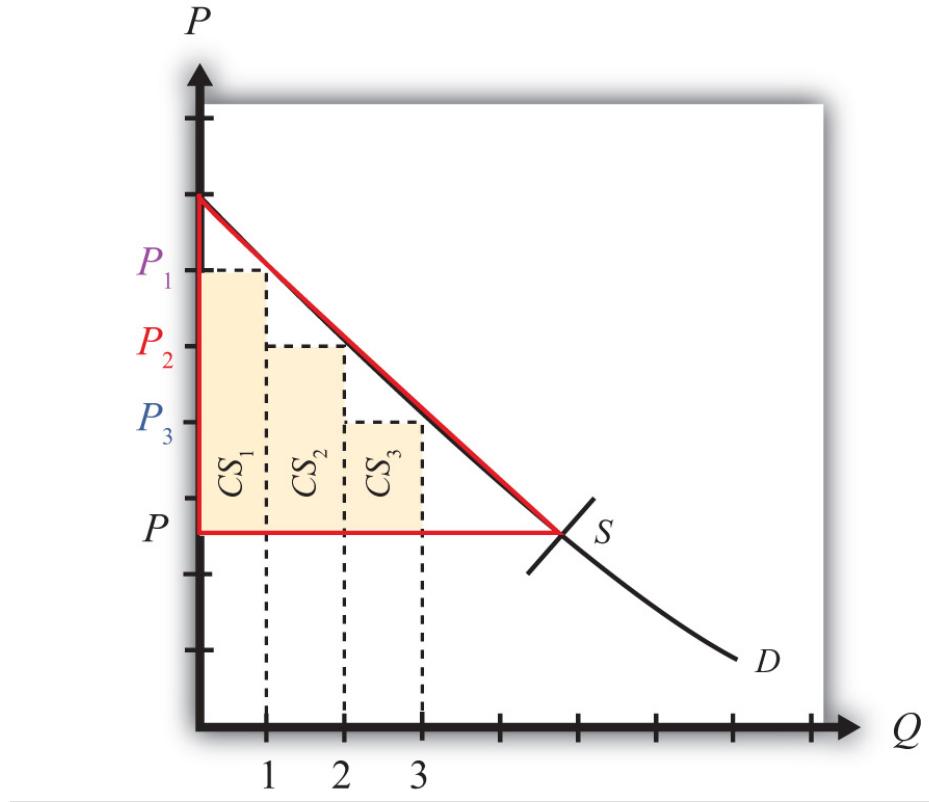
Consumer Surplus

Consumer surplus is used to measure the welfare of a group of consumers who purchase a particular product at a particular price. **Consumer surplus**⁶ is defined as the difference between what consumers are willing to pay for a unit of the good and the amount consumers actually do pay for the product. Willingness to pay can be read from a market demand curve for a product. The market demand curve shows the quantity of the good that would be demanded by all consumers at each and every price that might prevail. Read the other way, the demand curve tells us the maximum price that consumers would be willing to pay for any quantity supplied to the market.

A graphical representation of consumer surplus can be derived by considering the following exercise. Suppose that only one unit of a good is available in a market. As shown in [Figure 7.7 "Calculating Consumer Surplus"](#), that first unit could be sold at the price P_1 . In other words, there is a consumer in the market who would be willing to pay P_1 . Presumably that person either has a relatively high desire or need for the product or the person has a relatively high income. To sell two units of the good, the price would have to be lowered to P_2 . (This assumes that the firm cannot perfectly price discriminate and charge two separate prices to two customers.) A slightly lower price might induce another customer to purchase the product or might induce the first customer to buy two units. Three units of the good could be sold if the price is lowered to P_3 , and so on.

6. The difference between what consumers are willing to pay for a unit of the good and the amount consumers actually do pay for the product.

Figure 7.7 Calculating Consumer Surplus



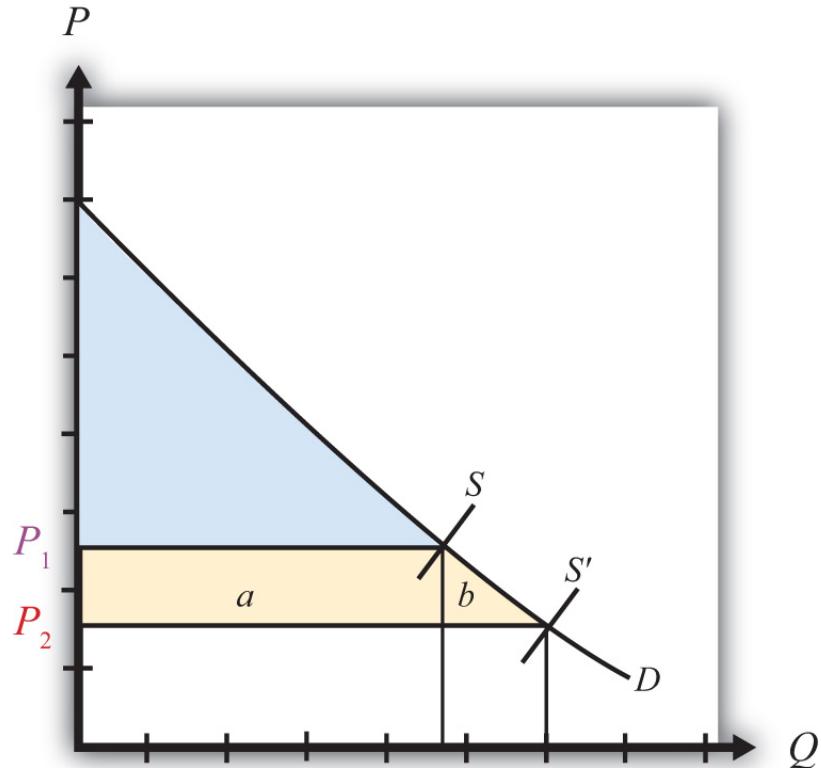
The price that ultimately prevails in a free market is that price that equalizes market supply with market demand. That price will be P in [Figure 7.7 "Calculating Consumer Surplus"](#) as long as the firms do not price discriminate. Now let's go back to the first unit that could have been sold. The person who would have been willing to pay P_1 for a unit of the good ultimately pays only P for the unit. The difference between the two prices represents the amount of consumer surplus that accrues to that person. For the second unit of the good, someone would have been willing to pay P_2 but ultimately pays P . The second unit generates a smaller amount of surplus than the first unit.

We can continue this procedure until the market supply at the price P is reached. The total consumer surplus in the market is given by the sum of the areas of the rectangles. If many units of the product are sold, then a one-unit width would be much smaller than shown in [Figure 7.7 "Calculating Consumer Surplus"](#). Thus total consumer surplus can reasonably be measured as the area between the demand curve and the horizontal line drawn at the equilibrium market price. This is shown as the red triangle in the diagram. The area representing consumer surplus is measured in dollars.

Changes in Consumer Surplus

Suppose the supply of a good rises, represented by a rightward shift in the supply curve from S to S' in [Figure 7.8 "Depicting a Change in Consumer Surplus"](#). At the original price, P_1 , consumer surplus is given by the blue area in the diagram (the triangular area between the P_1 price line and the demand curve). The increase in supply lowers the market price to P_2 . The new level of consumer surplus is now given by the sum of the blue and yellow areas in [Figure 7.8 "Depicting a Change in Consumer Surplus"](#) (the triangular area between the P_2 price line and the demand curve). The change in consumer surplus, CS , is given by the yellow area in [Figure 7.8 "Depicting a Change in Consumer Surplus"](#) (the area denoted by a and b). Note that the change in consumer surplus is determined as the area between the price that prevails before, the price that prevails after, and the demand curve. In this case, consumer surplus rises because the price falls. Two groups of consumers are affected. Consumers who would have purchased the product even at the higher price, P_1 , now receive more surplus ($P_1 - P_2$) for each unit they purchase. These extra benefits are represented by the rectangular area a in the diagram. Also, there are additional consumers who were unwilling to purchase the product at price P_1 but are now willing to purchase at the price P_2 . Their consumer surplus is given by the triangular area b in the diagram.

Figure 7.8 Depicting a Change in Consumer Surplus



Producer Surplus

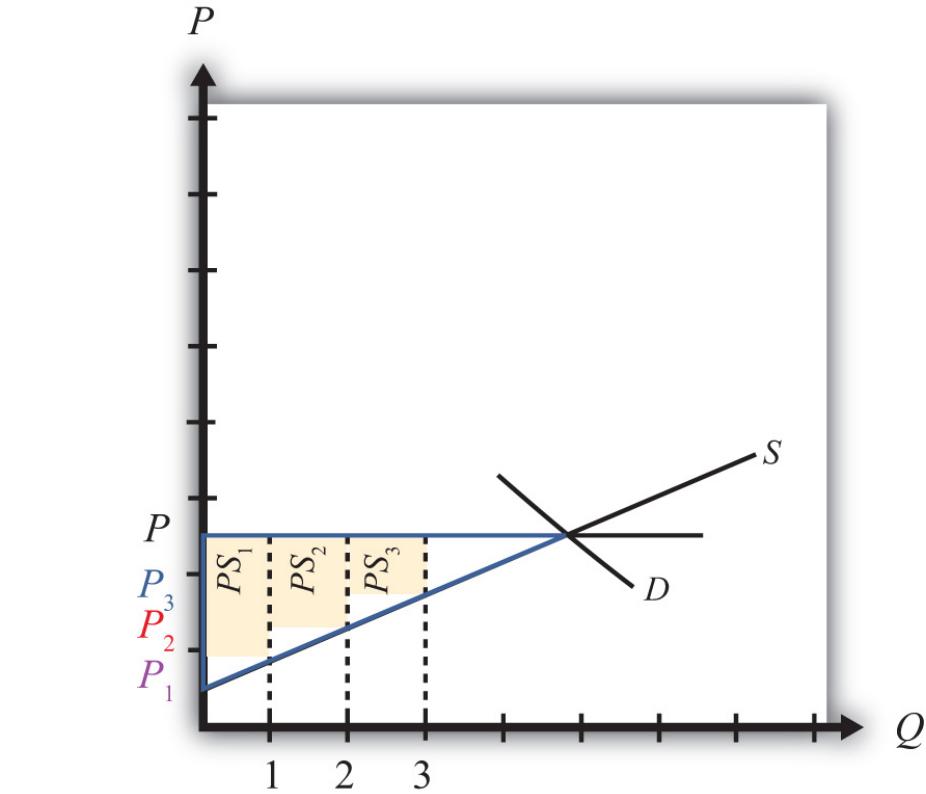
Producer surplus is used to measure the welfare of a group of firms that sell a particular product at a particular price. **Producer surplus**⁷ is defined as the difference between what producers actually receive when selling a product and the amount they would be willing to accept for a unit of the good. Firms' willingness to accept payments can be read from a market supply curve for a product. The market supply curve shows the quantity of the good that firms would supply at each and every price that might prevail. Read the other way, the supply curve tells us the minimum price that producers would be willing to accept for any quantity demanded by the market.

A graphical representation of producer surplus can be derived by considering the following exercise. Suppose that only one unit of a good is demanded in a market. As shown in [Figure 7.9 "Calculating Producer Surplus"](#), some firm would be willing to accept the price P_1 if only one unit is produced. If two units of the good were demanded in the market, then the minimum price to induce two units to be supplied is P_2 . A slightly higher price would induce another firm to supply an

7. The difference between what producers actually receive when selling a product and the amount they would be willing to accept for a unit of the good.

additional unit of the good. Three units of the good would be made available if the price were raised to P_3 , and so on.

Figure 7.9 *Calculating Producer Surplus*



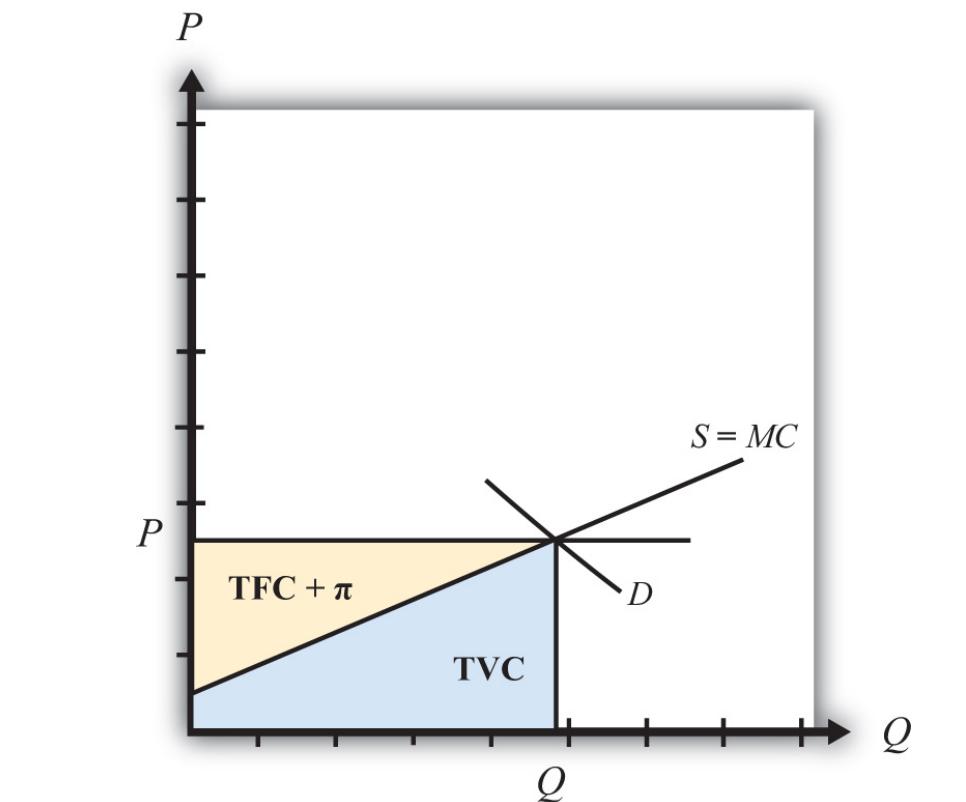
The price that ultimately prevails in a free market is the price that equalizes market supply with market demand. That price will be P in [Figure 7.9 "Calculating Producer Surplus"](#). Now let's go back to the first unit demanded. Some firm would have been willing to supply one unit at the price P_1 but ultimately receives P for the unit. The difference between the two prices represents the amount of producer surplus that accrues to the firm. For the second unit of the good, some firm would have been willing to supply the unit at the price P_2 but ultimately receives P . The second unit generates a smaller amount of surplus than the first unit.

We can continue this procedure until the market demand at the price P is reached. The total producer surplus in the market is given by the sum of the areas of the rectangles. If many units of the product are sold, then the one-unit width would be much smaller than shown in [Figure 7.9 "Calculating Producer Surplus"](#). Thus total producer surplus can reasonably be measured as the area between the supply curve and the horizontal line drawn at the equilibrium market price. This is shown as the

yellow triangle in the diagram. The area representing producer surplus is measured in dollars.

Producer surplus can be interpreted as the amount of revenue allocated to fixed costs and profit in the industry. This is because the market supply curve corresponds to industry marginal costs. Recall that firms choose output in a perfectly competitive market by setting the price equal to the marginal cost. Thus the marginal cost is equal to the price P in [Figure 7.10 "Interpreting Producer Surplus"](#) at an industry output equal to Q . The marginal cost represents the addition to cost for each additional unit of output. As such, it represents an additional *variable* cost for each additional unit of output. This implies that the area under the supply curve at an output level such as Q represents the total variable cost (TVC) to the industry, shown as the blue area in [Figure 7.10 "Interpreting Producer Surplus"](#).

Figure 7.10 Interpreting Producer Surplus



On the other hand, the market price multiplied by the quantity produced ($P \times Q$) represents the total revenue received by firms in the industry. This is represented by the sum of the blue and yellow areas in the diagram. The difference between the total revenue and the total variable cost, in turn, represents payments made to fixed factors of production, or total fixed cost (TFC), and any short-run profits (Π).

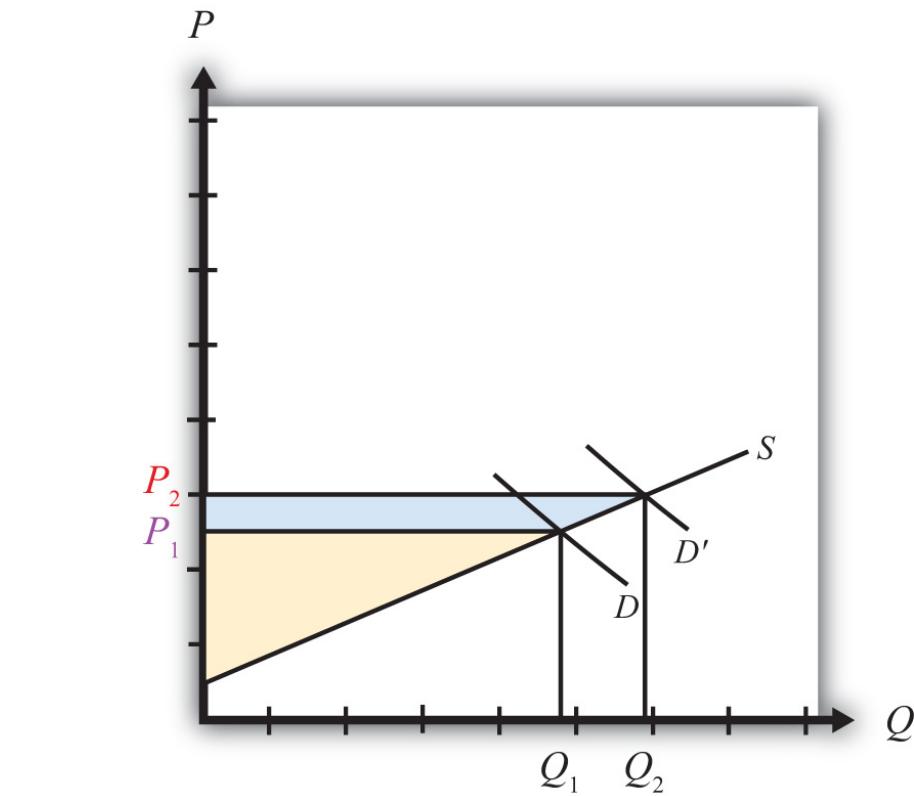
accruing to firms in the industry (the yellow area in the figure—that is, the area between the price line and the supply curve). This area is the same as the producer surplus.

Since fixed factors of production represent capital equipment that must be installed by the owners of the firms before any output can be produced, it is reasonable to use producer surplus to measure the well-being of the owners of the firms in the industry.

Changes in Producer Surplus

Suppose the demand for a good rises, represented by a rightward shift in the demand curve from D to D' in [Figure 7.11 "Depicting a Change in Producer Surplus"](#). At the original price, P_1 , producer surplus is given by the yellow area in [Figure 7.11 "Depicting a Change in Producer Surplus"](#) (the triangular area between the P_1 price and the supply curve). The increase in demand raises the market price to P_2 . The new level of producer surplus is now given by the sum of the blue and yellow areas in the figure (the triangular area between the price P_2 and the supply curve). The change in producer surplus, PS , is given by the blue area in the figure (the area between the two prices and the supply curve). Note that the change in producer surplus is determined as the area between the price that prevails before, the price that prevails after, and the supply curve. In this case, producer surplus rises because the price increases and output rises. The increase in price and output raises the return to fixed costs and the profitability of firms in the industry. The increase in output also requires an increase in variable factors of production such as labor. Thus one additional benefit to firms not measured by the increase in producer surplus is an increase in industry employment.

Figure 7.11 Depicting a Change in Producer Surplus



KEY TAKEAWAYS

- Consumer surplus and producer surplus are methods used to identify the magnitude of the welfare effects on consumers of a product and producers of a product.
- Consumer surplus measures the extra amount of money consumers would be willing to pay for a product over what they actually did pay.
- Consumer surplus is measured as the area between the demand curve, the horizontal line at the equilibrium price, and the vertical axis.
- Producer surplus is the extra amount of money producers receive when selling a product above what they would be willing to accept for it.
- Producer surplus is measured as the area between the supply curve, the horizontal line at the equilibrium price, and the vertical axis.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a measure of consumer welfare in a partial equilibrium analysis.
 - b. The term used to describe a measure of producer welfare in a partial equilibrium analysis.
 - c. Of *increase, decrease, or stay the same*, this is the effect of a price decrease on consumer surplus.
 - d. Of *increase, decrease, or stay the same*, this is the effect of a price increase on producer surplus.
 - e. Of *increase, decrease, or stay the same*, this is the effect of a demand increase on producer surplus.
 - f. Of *increase, decrease, or stay the same*, this is the effect of a supply increase on consumer surplus.
2. Suppose the demand for baseballs is given by $D = 1,000 - 20P$.
 - a. Calculate consumer surplus at a market price of \$20.
 - b. Calculate the change in consumer surplus if the price increases by \$5.
3. Suppose the supply of baseballs is given by $S = 30P$.
 - a. Calculate producer surplus at a market price of \$20.
 - b. Calculate the change in producer surplus if the price decreases by \$5.

7.4 Import Tariffs: Large Country Price Effects

LEARNING OBJECTIVES

1. Identify the effects of a specific tariff on prices in both countries and the quantity traded.
2. Know the equilibrium conditions that must prevail in a tariff equilibrium.

Suppose Mexico, the importing country in free trade, imposes a specific tariff on imports of wheat. As a tax on imports, the tariff will inhibit the flow of wheat across the border. It will now cost more to move the product from the United States into Mexico.

As a result, the supply of wheat to the Mexican market will fall, inducing an increase in the price of wheat. Since wheat is homogeneous and the market is perfectly competitive, the price of all wheat sold in Mexico, both Mexican wheat and U.S. imports, will rise in price. The higher price will reduce Mexico's import demand.

The reduced wheat supply to Mexico will shift back supply to the U.S. market. Since Mexico is assumed to be a large importer, the supply shifted back to the U.S. market will be enough to induce a reduction in the U.S. price. The lower price will reduce the U.S. export supply.

For this reason, a country that is a large importer is said to have **monopsony power in trade**⁸. A monopsony arises whenever there is a single buyer of a product. A monopsony can gain an advantage for itself by reducing its demand for a product in order to induce a reduction in the price. In a similar way, a country with monopsony power can reduce its demand for imports (by setting a tariff) to lower the price it pays for the imported product. Note that these price effects are identical in direction to the price effects of an import quota, a voluntary export restraint, and an export tax.

8. Another term to describe a large importing country—that is, a country whose policy actions can affect international prices.

A new tariff-ridden equilibrium will be reached when the following two conditions are satisfied:

$$P_T^{Mex} = P_T^{US} + T$$

and

$$XS^{US}(P_T^{US}) = MD^{Mex}(P_T^{Mex}),$$

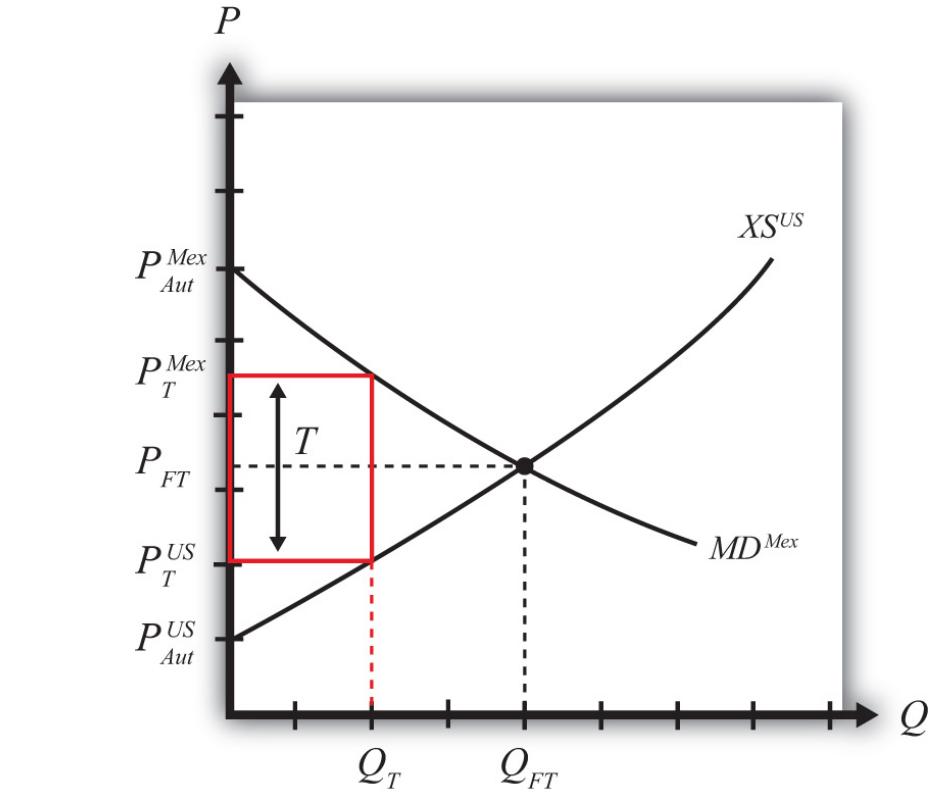
where T is the tariff, P_T^{Mex} is the price in Mexico after the tariff, and P_T^{US} is the price in the United States after the tariff.

The first condition represents a price wedge between the final U.S. price and the Mexican price equal to the amount of the tariff. The prices must differ by the tariff because U.S. suppliers of wheat must receive the same price for their product regardless of whether the product is sold in the United States or Mexico, and all wheat sold in Mexico must be sold at the same price. Since a tax is collected at the border, the only way for these price equalities within countries to arise is if the price differs across countries by the amount of the tax.

The second condition states that the amount the United States wants to export at its new lower price must be equal to the amount Mexico wants to import at its new higher price. This condition guarantees that world supply of wheat equals world demand for wheat.

The tariff equilibrium is depicted graphically in [Figure 7.12 "Depicting a Tariff Equilibrium: Large Country Case"](#). The Mexican price of wheat rises from P_{FT} to P_T^{Mex} , which reduces its import demand from Q_{FT} to Q_T . The U.S. price of wheat falls from P_{FT} to P_T^{US} , which also reduces its export supply from Q_{FT} to Q_T . The difference in the prices between the two markets is equal to the specific tariff rate, T .

Figure 7.12 Depicting a Tariff Equilibrium: Large Country Case



Notice that there is a unique set of prices that satisfies the equilibrium conditions for every potential tariff that is set. If the tariff were set higher than T , the price wedge would rise, causing a further increase in the Mexican price, a further decrease in the U.S. price, and a further reduction in the quantity traded.

At the extreme, if the tariff were set equal to the difference in autarky prices (i.e., $T = P_{Aut}^{Mex} - P_{Aut}^{US}$), then the quantity traded would fall to zero. In other words, the tariff would *prohibit* trade. Indeed, any tariff set greater than or equal to the difference in autarky prices would eliminate trade and cause the countries to revert to autarky in that market. Thus we define a prohibitive tariff as any tariff, T_{pro} , such that

$$T_{pro} \geq P_{Aut}^{Mex} - P_{Aut}^{US}.$$

The Price Effects of a Tariff: A Simple Dynamic Story

For an intuitive explanation about why these price changes would likely occur in a real-world setting, read the following story about the likely dynamic adjustment

process. Technically, this story is *not* a part of the partial equilibrium model, which is a static model that does not contain adjustment dynamics. However, it is worthwhile to think about how a real market adjusts to the equilibria described in these simple models.

Suppose the United States and Mexico are initially in a free trade equilibrium. Mexico imports wheat at the free trade price of \$10 per bushel. Imagine that the market for unprocessed wheat in both the United States and Mexico is located in a warehouse in each country. Each morning, wheat arrives from the suppliers and is placed in the warehouse for sale. During the day, consumers of unprocessed wheat arrive to buy the supply. For simplicity, assume there is no service charge collected by the intermediary that runs the warehouses. Thus, for each bushel sold, \$10 passes from the consumer directly to the producer.

Each day, the wheat market clears in the United States and Mexico at the price of \$10. This means that the quantity of wheat supplied at the beginning of the day is equal to the quantity purchased by consumers during the day. Supply equals demand in each market at the free trade price of \$10.

Now suppose that Mexico places a \$2 specific tariff on imports of wheat. Let's assume that the agents in the model react slowly and rather naively to the change. Let's also suppose that the \$2 tariff is a complete surprise.

Each day, prior to the tariff, trucks carrying U.S. wheat would cross the Mexican border in the wee hours, unencumbered, en route to the Mexican wheat market. On the day the tariff is imposed, the trucks are stopped and inspected. The drivers are informed that they must pay \$2 for each bushel that crosses into Mexico.

Suppose the U.S. exporters of wheat naively pay the tax and ship the same number of bushels to the Mexican market that day. However, to recoup their losses, they raise the price by the full \$2. The wheat for sale in Mexico now is separated into two groups. The imported U.S. wheat now has a price tag of \$12, while the Mexican-supplied wheat retains the \$10 price. Mexican consumers now face a choice. However, since Mexican and U.S. wheat are homogeneous, the choice is simple. Every Mexican consumer will want to purchase the Mexican wheat at \$10. No one will want the U.S. wheat. Of course, sometime during the day, Mexican wheat will run out and consumers will either have to buy the more expensive wheat or wait till the next day. Thus some \$12 U.S. wheat will sell, but not the full amount supplied. At the end of the day, a surplus will remain. This means that there will be an excess demand for Mexican wheat and an excess supply of U.S. wheat in the Mexican market.

Mexican producers of wheat will quickly realize that they can supply more to the market and raise their price. A higher price is possible because the competition is now charging \$12. The higher supply and higher price will raise the profitability of the domestic wheat producers. (Note that the supply of wheat may not rise quickly since it is grown over an annual cycle. However, the supply of a different type of good could be raised rapidly. The length of this adjustment will depend on the nature of the product.) U.S. exporters will quickly realize that no one wants to buy their wheat at a price of \$12. Their response will be to reduce export supply and lower their price in the Mexican market.

As time passes, in the Mexican market, the price of Mexican-supplied wheat will rise from \$10 and the price of U.S. supplied wheat will fall from \$12 until the two prices meet somewhere in between. The homogeneity of the goods requires that if both goods are to be sold in the Mexican market, then they must sell at the same price in equilibrium.

As these changes take place in the Mexican market, other changes occur in the U.S. market. When U.S. exporters of wheat begin to sell less in Mexico, that excess supply is shifted back to the U.S. market. The warehouse in the United States begins to fill up with more wheat than U.S. consumers are willing to buy at the initial price of \$10. Thus at the end of each day, wheat supplies remain unsold. An inventory begins to pile up. Producers realize that the only way to unload the excess wheat is to cut the price. Thus the price falls in the U.S. market. At lower prices, though, U.S. producers are willing to supply less, thus production is cut back as well.

In the end, the U.S. price falls and the Mexican price rises until the two prices differ by \$2, the amount of the tariff. A Mexican price of \$11.50 and a U.S. price of \$9.50 is one possibility. A Mexican price of \$11 and a U.S. price of \$9 is another. U.S. producers now receive the same lower price for wheat whether they sell in the United States or Mexico. The exported wheat is sold at the higher Mexican price, but \$2 per bushel is paid to the Mexican government as tariff revenue. Thus U.S. exporters receive the U.S. price for the wheat sold in Mexico.

The higher price in Mexico raises domestic supply and reduces domestic demand, thus reducing their demand for imports. The lower price in the United States reduces U.S. supply, raises U.S. demand, and thus lowers U.S. export supply to Mexico. In a two-country world, the \$2 price differential that arises must be such that U.S. export supply equals Mexican import demand.

Noteworthy Price Effects of a Tariff

Two of the effects of a tariff are worthy of emphasis. First, although a tariff represents a tax placed solely on imported goods, the domestic price of both imported and domestically produced goods will rise. In other words, a tariff will cause local producers of the product to raise their prices. Why?

In the model, it is assumed that domestic goods are perfectly substitutable for imported goods (i.e., the goods are homogeneous). When the price of imported goods rises due to the tariff, consumers will shift their demand from foreign to domestic suppliers. The extra demand will allow domestic producers an opportunity to raise output and prices to clear the market. In so doing, they will also raise their profit. Thus as long as domestic goods are substitutable for imports and as long as the domestic firms are profit seekers, the price of the domestically produced goods will rise along with the import price.

The average consumer may not recognize this rather obvious point. For example, suppose the United States places a tariff on imported automobiles. Consumers of U.S.-made automobiles may fail to realize that they are likely to be affected. After all, they might reason, the tax is placed only on imported automobiles. Surely this would raise the imports' prices and hurt consumers of foreign cars, but why would that affect the price of U.S. cars? The reason, of course, is that the import car market and the domestic car market are interconnected. Indeed, the only way U.S.-made car prices would not be affected by the tariff is if consumers were completely unwilling to substitute U.S. cars for imported cars or if U.S. automakers were unwilling to take advantage of a profit-raising possibility. These conditions are probably unlikely in most markets around the world.

The second interesting price effect arises because the importing country is large. When a large importing country places a tariff on an imported product, it will cause the *foreign price to fall*. The reason? The tariff will reduce imports into the domestic country, and since its imports represent a sizeable proportion of the world market, world demand for the product will fall. The reduction in demand will force profit-seeking firms in the rest of the world to lower output and price in order to clear the market.

The effect on the foreign price is sometimes called the terms of trade effect. The terms of trade is sometimes defined as the price of a country's *export goods* divided by the price of its *import goods*. Here, since the importing country's import good will fall in price, the country's terms of trade will rise. Thus a tariff implemented by a large country will cause an improvement in the country's terms of trade.

KEY TAKEAWAYS

- An import tariff will raise the domestic price and, in the case of a large country, lower the foreign price.
- An import tariff will reduce the quantity of imports.
- An import tariff will raise the price of the “untaxed” domestic import-competing good.
- The tariff will drive a price wedge, equal to the tariff value, between the foreign price and the domestic price of the product.
- With the tariff in place in a two-country model, export supply at the lower foreign price will equal import demand at the higher domestic price.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The kind of power a country is said to have when its imports make up a significant share of the world market.
 - b. The direction of change of the domestic price after an import tariff is implemented by a domestic country.
 - c. The direction of change of the foreign price after an import tariff is implemented by a large domestic country.
 - d. The term used to describe a tariff that eliminates trade.
 - e. Of *increase, decrease, or stay the same*, this is the effect on the price of U.S.-made automobiles if the United States places a tax on imported foreign automobiles.
 - f. The price of tea in the exporting country if the importer sets a tariff of \$1.50 per pound and if the importer country price is \$5.50 inclusive of the tariff.
 - g. Of *increase, decrease, or stay the same*, this is the effect on imports of wheat if a wheat tariff is implemented.
 - h. Of *increase, decrease, or stay the same*, this is the effect on foreign exports of wheat if a wheat tariff is implemented by an importing country.
2. Complete the following descriptions of the equilibrium conditions with a tariff in place.
 - a. _____ is equal to the price in the exporting market with the foreign tariff *plus* the tariff.
 - b. Import demand, at the price that prevails in the importing country after the tariff, is *equal to* _____ at the price that prevails _____.

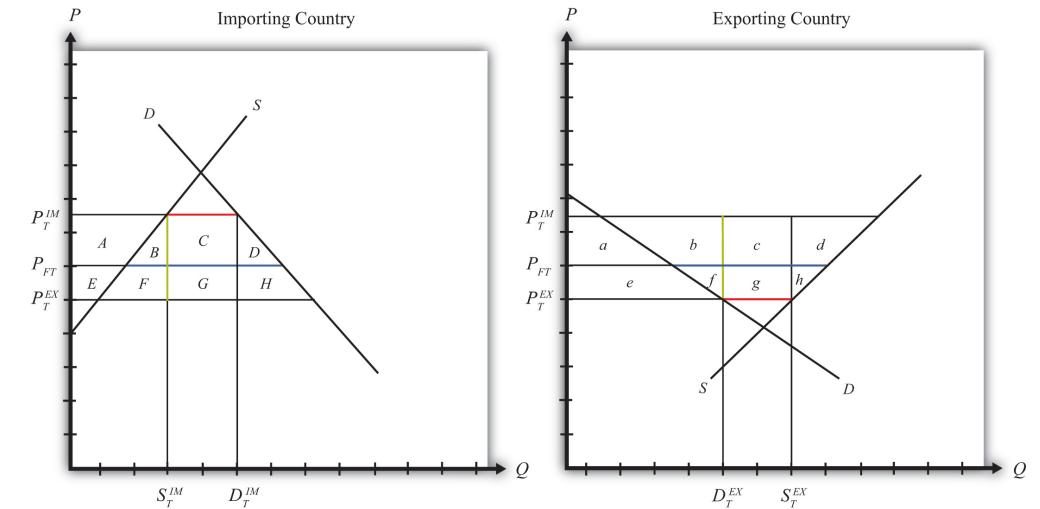
7.5 Import Tariffs: Large Country Welfare Effects

LEARNING OBJECTIVES

1. Use a partial equilibrium diagram to identify the welfare effects of an import tariff on producer and consumer groups and the government in the importing and exporting countries.
2. Calculate the national and world welfare effects of an import tariff.

Suppose that there are only two trading countries: one importing country and one exporting country. The supply and demand curves for the two countries are shown in [Figure 7.13 "Welfare Effects of a Tariff: Large Country Case"](#). P_{FT} is the free trade equilibrium price. At that price, the excess demand by the importing country equals excess supply by the exporter.

Figure 7.13 Welfare Effects of a Tariff: Large Country Case



The quantity of imports and exports is shown as the blue line segment on each country's graph. (That's the horizontal distance between the supply and demand curves at the free trade price.) When a large importing country implements a tariff it will cause an increase in the price of the good on the domestic market and a decrease in the price in the rest of the world (RoW). Suppose after the tariff the price in the importing country rises to P_T^{IM} and the price in the exporting country falls to P_T^{EX} . If the tariff is a specific tax, then the tariff rate would be $T = P_T^{IM} - P_T^{EX}$, equal to the length of the green line segment in the diagram. If

the tariff were an ad valorem tax, then the tariff rate would be given by

$$T = \frac{P_T^M}{P_T^{EX}} - 1.$$

Table 7.1 "Welfare Effects of an Import Tariff" provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the governments in the importing and exporting countries. The aggregate national welfare effects and the world welfare effects are also shown.

Table 7.1 Welfare Effects of an Import Tariff

	Importing Country	Exporting Country
Consumer Surplus	- (A + B + C + D)	+ e
Producer Surplus	+ A	- (e + f + g + h)
Govt. Revenue	+ (C + G)	0
National Welfare	+ G - (B + D)	- (f + g + h)
World Welfare	-; (B + D) - (f + h)	

Refer to Table 7.1 "Welfare Effects of an Import Tariff" and Figure 7.13 "Welfare Effects of a Tariff: Large Country Case" to see how the magnitudes of the changes are represented.

Tariff effects on the importing country's consumers. Consumers of the product in the importing country suffer a reduction in well-being as a result of the tariff. The increase in the domestic price of both imported goods and the domestic substitutes reduces the amount of consumer surplus in the market.

Tariff effects on the importing country's producers. Producers in the importing country experience an increase in well-being as a result of the tariff. The increase in the price of their product on the domestic market increases producer surplus in the industry. The price increases also induce an increase in the output of existing firms (and perhaps the addition of new firms); an increase in employment; and an increase in profit, payments, or both to fixed costs.

Tariff effects on the importing country's government. The government receives tariff revenue as a result of the tariff. Who benefits from the revenue depends on how the government spends it. Typically, the revenue is simply included as part of the general funds collected by the government from various sources. In this case, it is

impossible to identify precisely who benefits. However, these funds help support many government spending programs, which presumably help either most people in the country, as is the case with public goods, or certain worthy groups. Thus someone within the country is the likely recipient of these benefits.

Tariff effects on the importing country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the government. The net effect consists of three components: a positive terms of trade effect (G), a negative production distortion (B), and a negative consumption distortion (D).

Because there are both positive and negative elements, the net national welfare effect can be either positive or negative. The interesting result, however, is that it can be *positive*. This means that a tariff implemented by a large importing country *may* raise national welfare.

Generally speaking, the following are true:

1. Whenever a large country implements a small tariff, it will raise national welfare.
2. If the tariff is set too high, national welfare will fall.
3. There will be a positive optimal tariff that will maximize national welfare.

However, it is also important to note that not everyone's welfare rises when there is an increase in national welfare. Instead, there is a redistribution of income. Producers of the product and recipients of government spending will benefit, but consumers will lose. A national welfare increase, then, means that the sum of the gains exceeds the sum of the losses across all individuals in the economy. Economists generally argue that, in this case, compensation from winners to losers can potentially alleviate the redistribution problem.

Tariff effects on the exporting country's consumers. Consumers of the product in the exporting country experience an increase in well-being as a result of the tariff. The decrease in their domestic price raises the amount of consumer surplus in the market.

Tariff effects on the exporting country's producers. Producers in the exporting country experience a decrease in well-being as a result of the tariff. The decrease in the price of their product in their own market decreases producer surplus in the

industry. The price decline also induces a decrease in output, a decrease in employment, and a decrease in profit, payments, or both to fixed costs.

Tariff effects on the exporting country's government. There is no effect on the exporting country's government revenue as a result of the importer's tariff.

Tariff effects on the exporting country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers and producers. The net effect consists of three components: a negative terms of trade effect (g), a negative consumption distortion (f), and a negative production distortion (h).

Since all three components are negative, the importer's tariff must result in a reduction in national welfare for the exporting country. However, it is important to note that a redistribution of income occurs—that is, some groups gain while others lose. In this case, the sum of the losses exceeds the sum of the gains.

Tariff effects on world welfare. The effect on world welfare is found by summing the national welfare effects on the importing and exporting countries. By noting that the terms of trade gain to the importer is equal to the terms of trade loss to the exporter, the world welfare effect reduces to four components: the importer's negative production distortion (B), the importer's negative consumption distortion (D), the exporter's negative consumption distortion (f), and the exporter's negative production distortion (h). Since each of these is negative, the world welfare effect of the import tariff is negative. The sum of the losses in the world exceeds the sum of the gains. In other words, we can say that an import tariff results in a reduction in world production and consumption efficiency.

KEY TAKEAWAYS

- An import tariff lowers consumer surplus in the import market and raises it in the export country market.
- An import tariff raises producer surplus in the import market and lowers it in the export country market.
- The national welfare effect of an import tariff is evaluated as the sum of the producer and consumer surplus and government revenue effects.
- National welfare may rise or fall when a large country implements an import tariff.
- National welfare in the exporting country falls when an importing country implements an import tariff.
- An import tariff of any size will reduce world production and consumption efficiency and thus cause world welfare to fall.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The product of the specific tariff rate and the quantity of imports.
 - b. Of *increase, decrease, or stay the same*, this is the effect of a tariff on the welfare of consumers of the product in the large importing country.
 - c. Of *increase, decrease, or stay the same*, this is the effect of a tariff on the welfare of producers of the product in the large importing country.
 - d. Of *increase, decrease, or stay the same*, this is the effect of a tariff on the welfare of the recipients of government benefits in the large importing country.
 - e. Of *increase, decrease, or stay the same*, this is the effect of a tariff on the welfare of consumers of the product in the large exporting country.
 - f. Of *increase, decrease, or stay the same*, this is the effect of a tariff on the welfare of producers of the product in the exporting country.
 - g. Of *increase, decrease, or stay the same*, this is the effect of a tariff on the world welfare.
 - h. Of *larger, smaller, or the same*, this is how the magnitude of the consumer losses compares with the magnitude of the producer gains in an importing country implementing a tariff.
 - i. Of *larger, smaller, or the same*, this is how the magnitude of the consumer gains compares with the magnitude of the producer losses in an exporting country affected by a foreign tariff.
2. Consider the following trade policy actions (each applied by the domestic country) listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of each policy on the variables listed in the first column. Use a partial equilibrium model to determine the answers and assume that the shapes of the supply and demand curves are “normal.”

Assume that none of the policies begin with or result in prohibitive trade policies. Also assume that none of the policies correct for market imperfections or distortions. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

For example, an import tariff applied by a large country will cause an increase in the domestic price of the import good; therefore, a + is placed in the first box of the table.

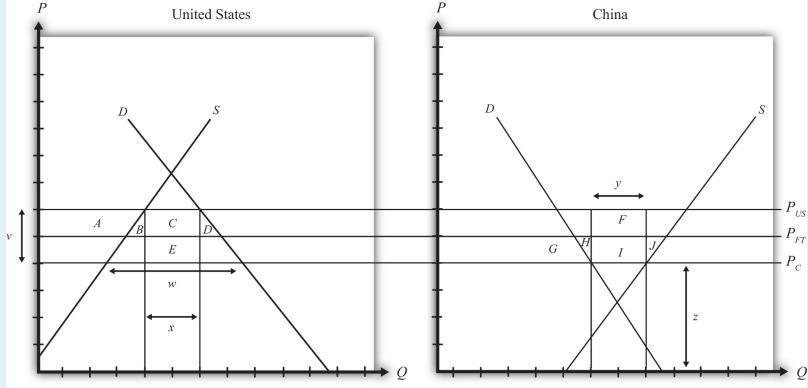
TABLE 7.2 TRADE POLICY EFFECTS

	I <i>Import Tariff by a Large Country—Initial Tariff Is Zero</i>	II <i>Import Tariff Reduction by a Large Country</i>
Domestic Market Price	+	
Domestic Industry Employment		
Domestic Consumer Welfare		

	I Import Tariff by a Large Country—Initial Tariff Is Zero	II Import Tariff Reduction by a Large Country
Domestic Producer Welfare		
Domestic Government Revenue		
Domestic National Welfare		
Foreign Price		
Foreign Consumer Welfare		
Foreign Producer Welfare		
Foreign National Welfare		

3. Consider the following partial equilibrium diagram depicting two countries, China and the United States, trading a product with each other. Suppose P_{FT} is the free trade price, P^{US} is the price in the United States when a tariff is in place, and P^C is the price in China when a tariff is in place. Answer the following questions by referring to the figure below. Assume the letters, A, B, C, D, E, F, G, H, I, and J refer to areas on the graph. The letters v, w, x, y, and z refer to lengths.

Figure 7.14
Two Large Trading Countries



- Which country is the exporter of the product?
- Where on the graph is the level of imports depicted with the tariff in place?
- Which areas on the graph represent the change in consumer surplus for the importing country if the tariff is removed? (Include the sign.)
- Which areas represent the tariff revenue lost by the importing government?
- Which areas represent the net national welfare effect of the tariff elimination by the importing country?
- Which areas represent the net national welfare effect of the tariff elimination in the exporting country?
- Which areas represent the world welfare effects of the tariff elimination?

7.6 The Optimal Tariff

LEARNING OBJECTIVES

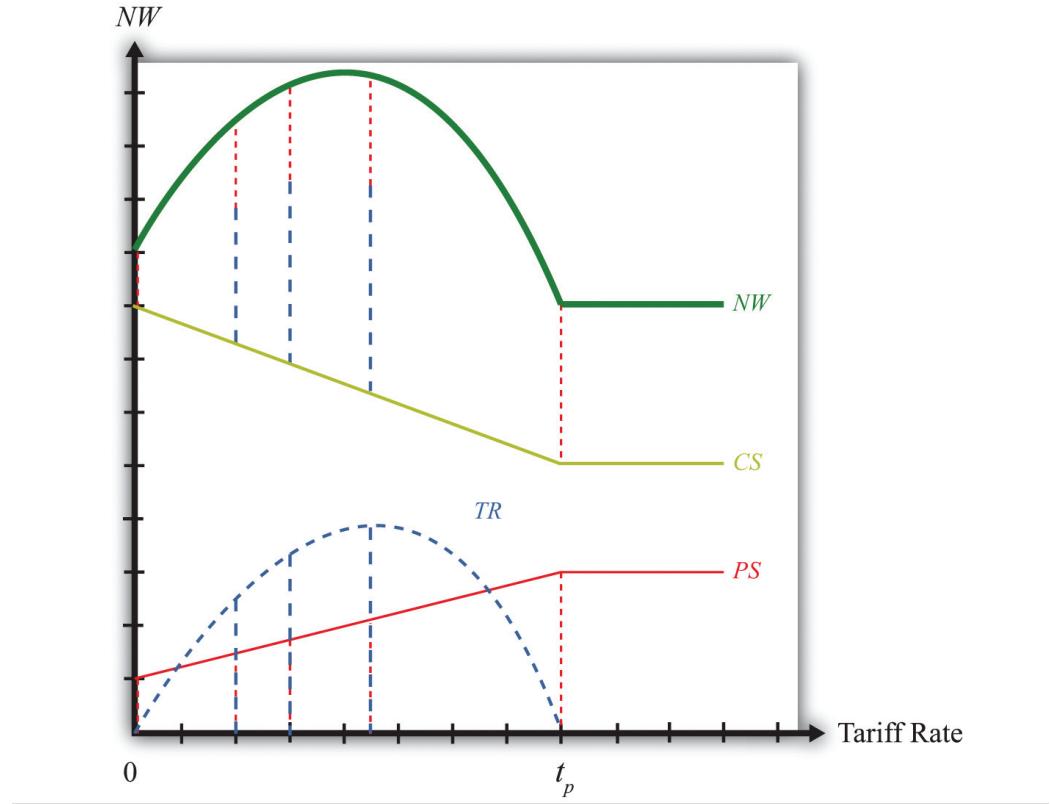
1. Plot the impact of an import tariff in a large country on consumer surplus, producer surplus, government revenue, and national welfare as the tariff is raised from zero.
2. Describe how tariff changes will affect national welfare in different circumstances.

The possibility that a tariff could improve national welfare for a large country in international markets was first noted by Robert Torrens. Since the welfare improvement occurs only if the terms of trade gain exceeds the total deadweight losses, the argument is commonly known as the *terms of trade argument* for protection.

Economists have studied the conditions under which a tariff will improve welfare in a variety of perfectly competitive models. This section describes the general results that come from that analysis.

Consider [Figure 7.15 "Derivation of the Optimal Tariff: Large Country"](#), which plots the levels of consumer surplus (CS), producer surplus (PS), and tariff revenue (TR) at different tariff rates. The origin corresponds to a zero tariff rate, or free trade. As the tariff is increased from zero, consumer surplus falls since the domestic price rises. This is shown by the solid declining (green) CS line. When the tariff becomes prohibitive at t_p , the price settles at the autarky price, and any further increases in the tariff have no effect on consumer surplus. Hence the CS line becomes flat above t_p .

Figure 7.15 Derivation of the Optimal Tariff: Large Country



Producer surplus (PS), the red dotted line, rises as the tariff is increased from zero; however, it rises at a lower rate than consumer surplus falls. This occurs because, for an importing country, producer surplus increases are less than the change in consumer surplus for any increase in the tariff. When the prohibitive tariff is reached, again the price settles at the autarky price, and any further increases in the tariff rate have no effect on producer surplus.

Tariff revenue (TR), the blue dashed line, first increases with the increase in the tariff and then decreases for higher tariff rates. This occurs because tariff revenue equals the tariff rate multiplied by imports. As the tariff is increased from zero, imports fall at a slower rate than the increase in the tariff rate, hence revenue rises. Eventually, imports begin to fall faster than the tariff rate rises, and tariff revenue declines. The tariff rate that generates the highest tariff revenue is called the **maximum revenue tariff**⁹.

Another way to see that tariff revenue must rise and then fall with increasing tariffs is to note that when the tariff rate is zero, tariff revenue has to be zero for any level of imports. Also, when the tariff rate is at or above t_p , the prohibitive tariff, imports

9. The tariff that achieves the highest government revenue.

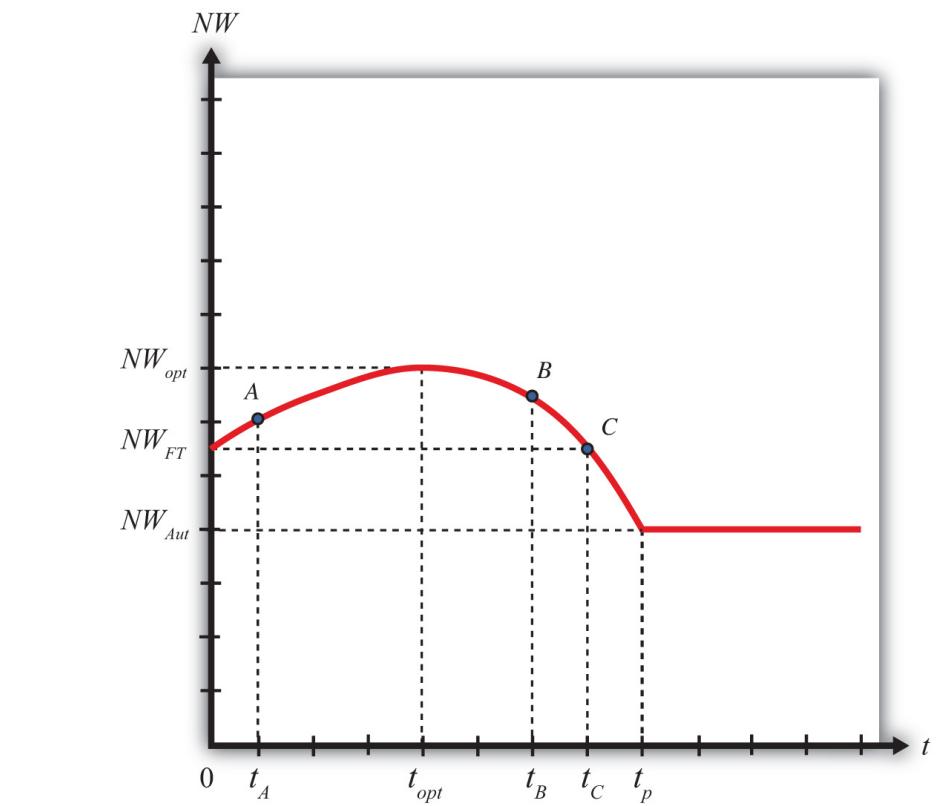
are zero, thus whatever the tariff rate, tariff revenue again must be zero.

Somewhere between a zero tariff and the prohibitive tariff, tariff revenue has to be positive. Thus tariff revenue must rise from zero and then fall back to zero when it reaches t_p .

The national welfare level at each tariff rate is defined as the sum of consumer surplus, producer surplus, and tariff revenue. The vertical summation of these three curves generates the national welfare (NW) curve given by the thick, solid blue-green line. In [Figure 7.15 "Derivation of the Optimal Tariff: Large Country"](#), the vertical summation is displayed for five different levels of the tariff rate.

The basic shape of the national welfare line is redrawn in [Figure 7.16 "Optimal Tariff: Large Country Case"](#). Note that national welfare first rises and then falls as the tariff is increased from zero. For one tariff rate (t_{opt}), the country can realize the highest level of national welfare (NW_{opt}), one that is higher than that achievable in free trade. We call that tariff rate the “optimal tariff.” One regularity that results is that the optimal tariff is always less than the maximum revenue tariff.

Figure 7.16 Optimal Tariff: Large Country Case



If the tariff is raised above the optimal rate, as with an increase from t_{opt} to t_B , then national welfare will fall. The terms of trade gain, which rises as low tariffs are increased, will begin to fall at a higher tariff rate. Since the deadweight losses continue to rise, both effects contribute to the decline in national welfare. Note, however, that at a tariff level like t_B , national welfare still exceeds the free trade level.

Eventually, at even higher tariff rates, national welfare will fall below the free trade level. In [Figure 7.16 "Optimal Tariff: Large Country Case"](#), this occurs at tariff rates greater than t_C . The higher the tariff is raised, the lower will be the level of imports. At a sufficiently high tariff, imports will be eliminated entirely. The tariff will prohibit trade. At the prohibitive tariff (t_p), there is no tariff revenue, which implies that the previously positive terms of trade gain is now zero. The only effect of the tariff is the deadweight loss. The economy is effectively in autarky, at least with respect to this one market, hence national welfare is at NW_{Aut} . Note that any additional increases in the tariff above t_p will maintain national welfare at NW_{Aut} since the market remains at the autarky equilibrium.

The National Welfare Effects of Trade Liberalization for a Large Country

Trade liberalization can be represented by a decrease in the tariff rate on imports into a country. If the country is large in international markets, then the analysis in this chapter suggests that the effect on national welfare will depend on the values of the original tariff rate and the liberalized tariff rate.

For example, if the tariff is reduced from t_{opt} to t_A , then national welfare will fall when the country liberalizes trade in this market. However, if the tariff is reduced from t_B to t_{opt} , then national welfare will rise when trade liberalization occurs. This implies that trade liberalization does not necessarily improve welfare for a large importing country.

KEY TAKEAWAYS

- The optimal tariff is positive for a large importing country.
- National welfare with a zero tariff (free trade) is always higher than national welfare with a prohibitive tariff.
- The maximum revenue tariff is larger than the optimal tariff.
- The reduction of a tariff by a large importing country will lower national welfare if the initial tariff is less than the optimal tariff.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. A term used to describe a tariff that will raise national welfare to the greatest extent for a large importing country.
 - b. The term used to describe the tariff rate that generates the largest amount of government revenue.
 - c. The tariff rate that corresponds to free trade.
 - d. The tariff rate that is just sufficient to eliminate trade with the rest of the world.
 - e. Of *higher, lower, or the same*, this is how national welfare in free trade compares with national welfare in autarky.
 - f. Of *higher, lower, or the same*, this is how national welfare at the optimal tariff compares with national welfare in autarky.
 - g. Of *higher, lower, or the same*, this is how national welfare at the maximum revenue tariff compares with national welfare at the optimal tariff.
 - h. Of *higher, lower, or the same*, this is how producer welfare in free trade compares with producer welfare in autarky.
 - i. Of *higher, lower, or the same*, this is how consumer welfare in free trade compares with consumer welfare in autarky.

7.7 Import Tariffs: Small Country Price Effects

LEARNING OBJECTIVES

1. Identify the effects of a specific tariff on prices in both countries and the quantity traded.
2. Know the equilibrium conditions that must prevail in a tariff equilibrium.

The small country assumption means that the country's imports are a very small share of the world market—so small that even a complete elimination of imports would have an imperceptible effect on world demand for the product and thus would not affect the world price. Thus when a tariff is implemented by a small country, there is no effect on the world price.

The small country assumption implies that the export supply curve is horizontal at the level of the world price. The small importing country takes the world price as exogenous since it can have no effect on it. The exporter is willing to supply as much of the product as the importer wants at the given world price.

When the tariff is placed on imports, two conditions must hold in the final equilibrium—the same two conditions as in the case of a large country—namely,

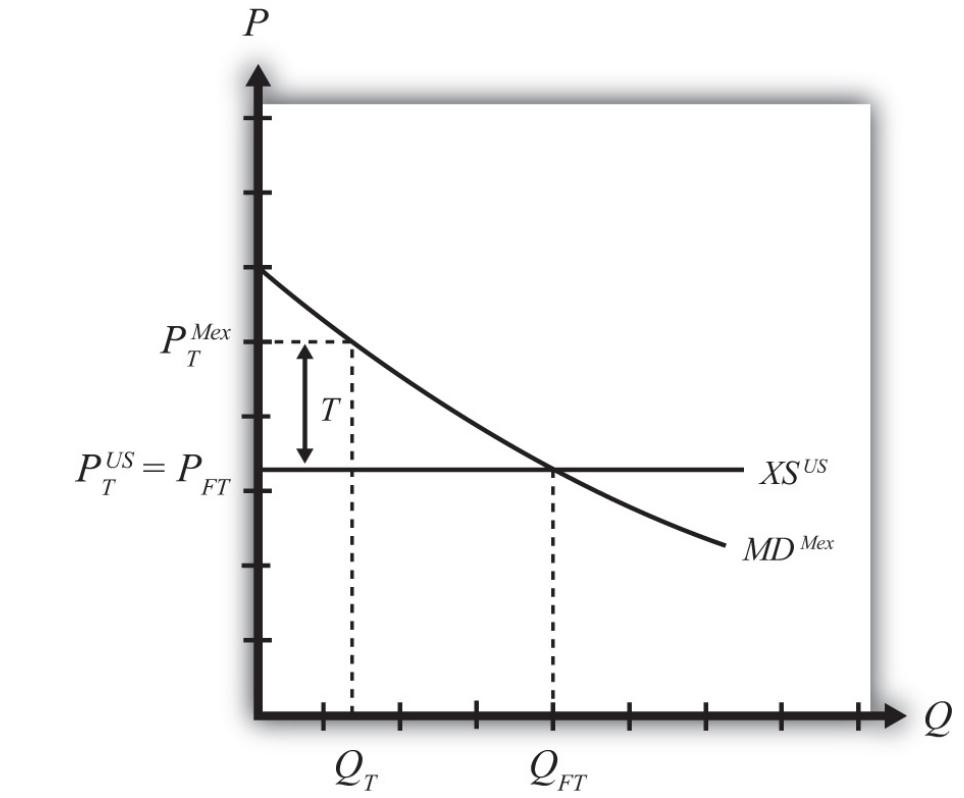
$$P_T^{Mex} = P_T^{US} + T$$

and

$$XS^{US}(P_T^{US}) = MD^{Mex}(P_T^{Mex}).$$

However, now P_T^{US} remains at the free trade price. This implies that, in the case of a small country, the price of the import good in the importing country will rise by the amount of the tariff, or in other words $P_T^{Mex} = P_{FT} + T$. As seen in [Figure 7.17 "Depicting a Tariff Equilibrium: Small Country Case"](#), the higher domestic price reduces import demand and export supply to Q_T .

Figure 7.17 Depicting a Tariff Equilibrium: Small Country Case



KEY TAKEAWAYS

- An import tariff will raise the domestic price and, in the case of a small country, leave the foreign price unchanged.
- An import tariff will reduce the quantity of imports.
- An import tariff will raise the domestic price of imports and import-competing goods by the full amount of the tariff.
- With the tariff in place in a two-country model, export supply at the unchanged foreign price will equal import demand at the higher domestic price.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The world price of butter if a small country has a tariff of \$0.50 per pound in place and butter sells for \$4.50 per pound.
 - b. The amount the domestic auto price rises if a small country places a \$100 tariff on auto imports.
 - c. Of *increase, decrease, or stay the same*, the effect on the world price when a small importing country implements a tariff.
 - d. Of *increase, decrease, or stay the same*, the effect on the import volume of a product when a small importing country implements a tariff.
 - e. Of *increase, decrease, or stay the same*, the effect on the exports from the rest of the world when a small importing country implements a tariff on the product.

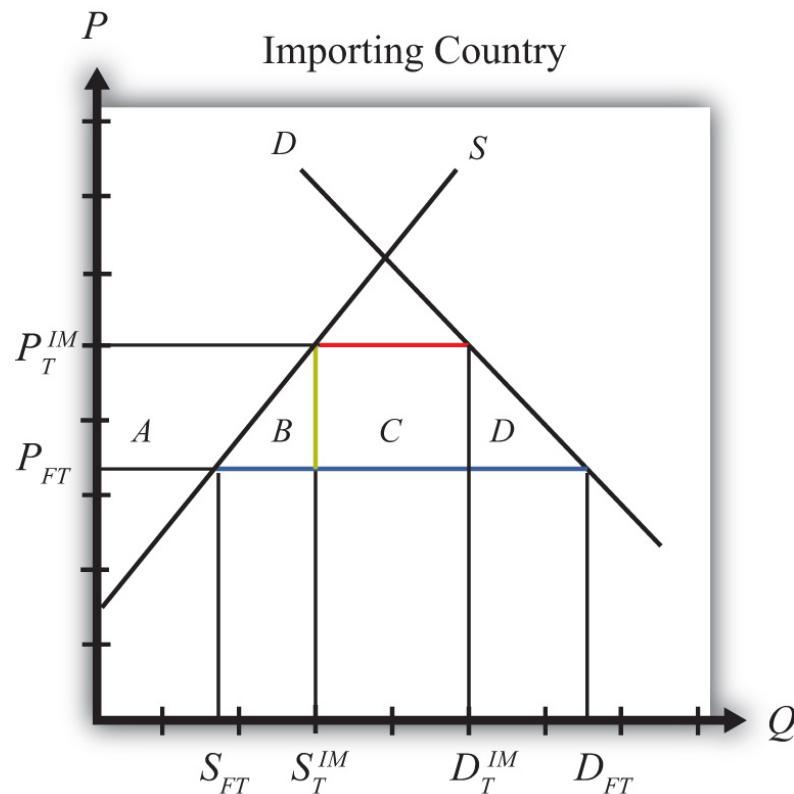
7.8 Import Tariffs: Small Country Welfare Effects

LEARNING OBJECTIVES

1. Use a partial equilibrium diagram to identify the welfare effects of an import tariff on producer and consumer groups and the government in the importing country.
2. Calculate the national welfare effects of an import tariff.

Consider a market in a small importing country that faces an international or world price of P_{FT} in free trade. The free trade equilibrium is depicted in [Figure 7.18 "Welfare Effects of a Tariff: Small Country Case"](#), where P_{FT} is the free trade equilibrium price. At that price, domestic demand is given by D_{FT} , domestic supply by S_{FT} , and imports by the difference $D_{FT} - S_{FT}$ (the blue line in the figure).

Figure 7.18 Welfare Effects of a Tariff: Small Country Case



When a specific tariff is implemented by a small country, it will raise the domestic price by the full value of the tariff. Suppose the price in the importing country rises to P_T^{IM} because of the tariff. In this case, the tariff rate would be $t = P_T^{IM} - P_{FT}$, equal to the length of the green line segment in the figure.

Table 7.3 "Welfare Effects of an Import Tariff" provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the governments in the importing country. The aggregate national welfare effect is also shown.

Table 7.3 Welfare Effects of an Import Tariff

	Importing Country
Consumer Surplus	- (A + B + C + D)
Producer Surplus	+ A
Govt. Revenue	+ C
National Welfare	- B - D

Refer to Table 7.3 "Welfare Effects of an Import Tariff" and Figure 7.18 "Welfare Effects of a Tariff: Small Country Case" to see how the magnitudes of the changes are represented.

Tariff effects on the importing country's consumers. Consumers of the product in the importing country are worse off as a result of the tariff. The increase in the domestic price of both imported goods and the domestic substitutes reduces consumer surplus in the market.

Tariff effects on the importing country's producers. Producers in the importing country are better off as a result of the tariff. The increase in the price of their product increases producer surplus in the industry. The price increases also induce an increase in the output of existing firms (and perhaps the addition of new firms), an increase in employment, and an increase in profit, payments, or both to fixed costs.

Tariff effects on the importing country's government. The government receives tariff revenue as a result of the tariff. Who will benefit from the revenue depends on how the government spends it. These funds help support diverse government spending programs; therefore, someone within the country will be the likely recipient of these benefits.

Tariff effects on the importing country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the government. The net effect consists of two components: a negative production efficiency loss (B) and a negative consumption efficiency loss (D). The two losses together are typically referred to as “deadweight losses.”

Because there are only negative elements in the national welfare change, the net national welfare effect of a tariff must be negative. This means that **a tariff implemented by a small importing country must reduce national welfare.**

In summary, the following are true:

1. Whenever a small country implements a tariff, national welfare falls.
2. The higher the tariff is set, the larger will be the loss in national welfare.
3. The tariff causes a redistribution of income. Producers and the recipients of government spending gain, while consumers lose.
4. Because the country is assumed to be small, the tariff has no effect on the price in the rest of the world; therefore, there are no welfare changes for producers or consumers there. Even though imports are reduced, the related reduction in exports by the rest of the world is assumed to be too small to have a noticeable impact.

KEY TAKEAWAYS

- An import tariff lowers consumer surplus and raises producer surplus in the import market.
- An import tariff by a small country has no effect on consumers, producers, or national welfare in the foreign country.
- The national welfare effect of an import tariff is evaluated as the sum of the producer and consumer surplus and government revenue effects.
- An import tariff of any size will result in deadweight losses and reduce production and consumption efficiency.
- National welfare falls when a small country implements an import tariff.

EXERCISES

1. Consider the following trade policy action (applied by the domestic country) listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of the policy on the variables listed in the first column. Use a partial equilibrium model to determine the answers, and assume that the shapes of the supply and demand curves are “normal.” Assume that the policy does not begin with, or result in, prohibitive trade policies. Also assume that the policy does not correct for market imperfections or distortions. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

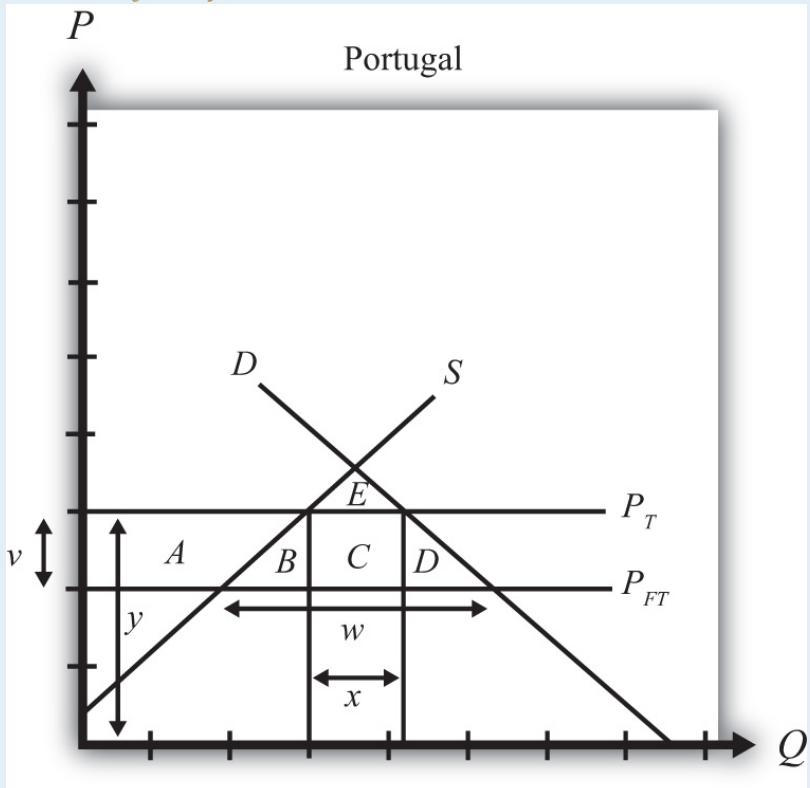
TABLE 7.4 TRADE POLICY EFFECTS

	Import Tariff Reduction by a Small Country
Domestic Market Price	
Domestic Industry Employment	
Domestic Consumer Welfare	
Domestic Producer Welfare	
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	

	Import Tariff Reduction by a Small Country
Foreign Consumer Welfare	
Foreign Producer Welfare	
Foreign National Welfare	

2. Consider the following partial equilibrium diagram depicting the market for radios in Portugal, a small importing country. Suppose P_{FT} is the free trade price and P_T is the price in Portugal when a tariff is in place. Answer the following questions by referring to the diagram. Assume the letters, A, B, C, D, and E refer to areas on the graph. The letters v, w, x, and y refer to lengths. (Be sure to include the direction of changes by indicating "+" or "-".)

Figure 7.19
A Small Trading Country



- a. Where on the graph is the level of imports in free trade?
- b. Which area or areas represent the *level* of consumer surplus in free trade?
- c. Which area or areas represent the *level* of producer surplus in free trade?
- d. Where on the graph is the size of the tariff depicted?
- e. Where on the graph is the level of imports after the tariff depicted?
- f. Which area or areas represent the tariff revenue collected by the importing government with the tariff in place?
- g. Which area or areas represent the change (+/-) in consumer surplus when the tariff is applied?
- h. Which area or areas represent the change (+/-) in producer surplus when the tariff is applied?
- i. Which area or areas represent the change (+/-) in national welfare when the tariff is applied?
- j. Which area or areas represent the efficiency losses that arise with the tariff?

7.9 Retaliation and Trade Wars

LEARNING OBJECTIVES

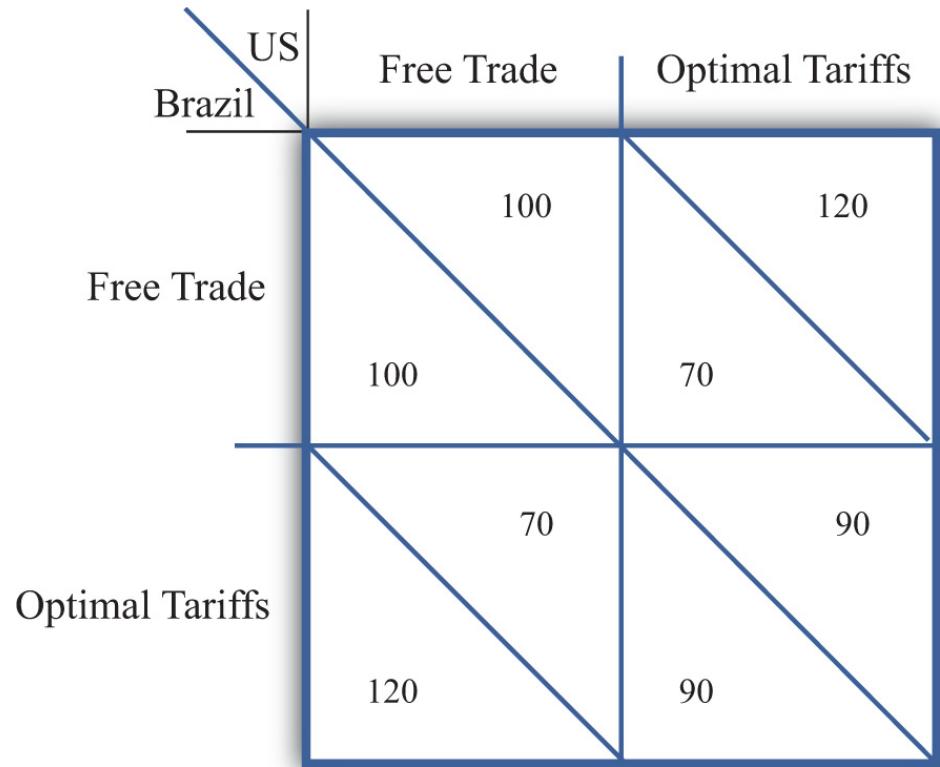
1. Identify the key components to describe an economic game, including players, strategies, objectives, and equilibrium concepts.
2. Determine both noncooperative and cooperative equilibria in an economic game.

The analysis of tariffs in a perfectly competitive market demonstrates that if a large country imposes a relatively small tariff, or if it imposes an optimal tariff, then domestic national welfare will rise but foreign national welfare will fall. The partial equilibrium analysis shows further that national welfare losses to the exporting nation exceed the national welfare gains to the importing nation. The reason is that any tariff set by a large country also reduces world welfare.

If we assume that nations are concerned about the national welfare effects of trade policies, then the tariff analysis provides a rationale for protectionism on the part of large importing nations. However, if large importing nations set optimal tariffs on all or many of their imported goods, the effect internationally will be to reduce the national welfare of its trading partners. If the trade partners are also concerned about their own national welfare, then they would likely find the optimal tariffs objectionable and would look for ways to mitigate the negative effects.

One effective way to mitigate the loss in national welfare, if the trade partners are also large countries, is to retaliate with optimal tariffs on your own imported goods. Thus if country A imports wine, cheese, and wheat from country B, and A places optimal tariffs on imports of these products, then country B could retaliate by imposing optimal tariffs on its imports of, say, lumber, televisions, and machine tools from country A. By doing so, country B could offset its national welfare losses in one set of markets with national welfare gains in another set.

Figure 7.20 A Trade Policy Game



We examine the effects of optimal tariffs and retaliation more formally by using a simple game theory setup. Suppose the players in the game are the governments of two large countries, the United States and Brazil. Suppose the United States imports a set of products (A, B, C , etc.) from Brazil, while Brazil imports a different set of products (X, Y, Z , etc.) from the United States. We imagine that each country's government must choose between two distinct trade policies, free trade and optimal tariffs. Each policy choice represents a game strategy. If the United States chooses free trade, then it imposes no tariffs on imports of goods A, B, C , and so on. If the United States chooses optimal tariffs, then it determines the optimal tariff in each import market and sets the tariff accordingly. Brazil is assumed to have the same set of policy choices available.

In Figure 7.20 "A Trade Policy Game", U.S. strategies are represented by the two columns; Brazilian strategies correspond to the two rows. The numbers represent the payoffs to the countries, measured as the level of national welfare realized in each country in each of the four possible scenarios. For example, if the United States chooses a free trade policy and Brazil chooses to impose optimal tariffs, then the payoffs are shown in the lower left-hand box. The Brazilian payoff is below the

diagonal, while the U.S. payoff is above the diagonal. Thus Brazil gets 120 units of welfare, while the United States gets 70 units.

Note that the size of the numbers used in the example is immaterial, but how they relate to the numbers in alternate boxes is not. We will use the results from the tariff analysis section to inform us about the relationship between the numbers.

To begin, let's assume that each country receives 100 units of national welfare when both the United States and Brazil choose free trade. If Brazil decides to impose optimal tariffs on all of its imports and the United States maintains its free trade position, then a partial equilibrium welfare analysis suggests the following:

1. Brazilian welfare will rise (we'll assume from 100 to 120 units).
2. U.S. welfare will fall (we'll assume from 100 to 70 units).
3. World welfare will fall (thus the sum of the U.S. and Brazilian welfare initially is 200 units but falls to $120 + 70 = 190$ afterward).

Similarly, if the United States imposes optimal tariffs on all of its imports while Brazil maintains free trade, then the countries will realize the payoffs in the upper right-hand box. The United States would get 120 units of welfare, while Brazil would get 70. To keep the example simple, we are assuming that the effects of tariffs are symmetric. In other words, the effect of U.S. optimal tariffs on the two countries is of the same magnitude as the effects of Brazilian tariffs.

Finally, if both countries set optimal tariffs against each other, then we can simply sum up the total effects. Since each country's actions raise its own welfare by 20 units and lower its trade partner's welfare by 30 units, when both countries impose tariffs, national welfare falls to 90 units in each country.

To determine which strategy the two governments would choose in this game, we need to identify the objectives of the players and the degree of cooperation. Initially, we will assume that each government is interested in maximizing its own national welfare and that the governments do not cooperate with each other. Afterward, we will consider the outcome when the governments do cooperate.

The Noncooperative Solution (Nash Equilibrium)

A noncooperative solution is a set of strategies such that each country maximizes its own national welfare subject to the strategy chosen by the other country. Thus, in general, if the U.S. strategy (r) maximizes U.S. welfare, when Brazil chooses its strategy (s) and if Brazil's strategy (s) maximizes Brazil's welfare when the United

States chooses strategy (r), then the strategy set (r,s) is a noncooperative solution to the game. A noncooperative solution is also commonly known as a Nash equilibrium.

How to Find a Nash Equilibrium

One can determine a Nash equilibrium in a simple two-player, two-strategy game by choosing a strategy for one of the players and answering the following series of questions:

1. Given the policy choice of the first player, what is the optimal policy of the second player?
2. Given the policy choice of the second player (from step one), what is the first player's optimal policy choice?
3. Given player one's optimal policy choice (from step two), what is the second player's optimal policy choice?

Continue this series of questions until neither player switches its strategy. Then this set of strategies is a Nash equilibrium.

In the trade policy game, the **Nash equilibrium**¹⁰ or noncooperative solution is the set of strategies (optimal tariffs, optimal tariffs). That is, both the United States and Brazil would choose to implement optimal tariffs. Why?

First, suppose the United States chooses the free trade strategy. Brazil's optimal policy, given the U.S. choice, is to implement optimal tariffs. This is because 120 units of national welfare are greater than 100 units. Second, if Brazil chooses optimal tariffs, then the optimal policy of the United States is optimal tariffs, since 90 units of welfare are greater than 70 units. Finally, if the United States chooses optimal tariffs, then Brazil's best choice is optimal tariffs since 90 is greater than 70.

The Cooperative Solution

A cooperative solution to a game is a set of strategies that would maximize the sum total of the benefits accruing to the players. In some instances, a cooperative outcome may require the transfer of goods or money between players to assure that each player is made better off than under alternative strategy choices. In this game, such a transfer is not required, however.

10. A game equilibrium in which every player is simultaneously maximizing his own profit given the choices being made by the other players.

The cooperative solution in the trade policy game is the set of strategies (free trade, free trade). At this outcome, total world welfare is at a maximum of 200 units.

Implications and Interpretations

First of all, notice that in the noncooperative game, each country is acting in its own best interests, yet the outcome is one that is clearly inferior for both countries relative to the cooperative strategy set (free trade, free trade). When both countries set optimal tariffs, each country realizes 90 units of welfare, while if both countries pursued free trade, each country would realize 100 units of welfare. This kind of result is often referred to as a prisoner's dilemma outcome. The dilemma is that pursuit of self-interest leads to an inferior outcome for both participants.

However, without cooperation, it may be difficult for the two countries to realize the superior free trade outcome. If both countries begin in free trade, each country has an individual incentive to deviate and implement optimal tariffs. And if either country does deviate, then the other would either suffer the welfare losses caused by the other country's restrictions or retaliate with tariff increases of its own in order to recoup some of the losses. This scenario in which one country retaliates in response to another's trade policy could be thought of as a trade war.

This story closely corresponds with events after the Smoot-Hawley Tariff Act was passed in the United States in 1930. The Smoot-Hawley Tariff Act raised tariffs to an average rate of 60 percent on many products imported into the United States. Although it is unlikely that the U.S. government set optimal tariffs, the tariffs nevertheless reduced foreign exports to the United States and injured foreign firms. In response to the U.S. tariffs, approximately sixty foreign nations retaliated and raised their tariffs on imports from the United States. The net effect was a substantial reduction in world trade, which very likely contributed to the length and severity of the Great Depression.

After World War II, the United States and other allied nations believed that high restrictions on trade were detrimental to growth in the world economy. The General Agreement on Tariffs and Trade (GATT) was initiated to promote trade liberalization among its member countries. The method of GATT was to hold multilateral tariff reduction "rounds." At each round, countries would agree to lower tariffs on imports by a certain average percentage in exchange for a reduction in tariffs by other countries by an equal percentage. Although GATT agreements never achieved a movement to free trade by all member countries, they do represent movements in that direction.

In a sense, then, the GATT represents an international cooperative agreement that facilitates movement toward the free trade strategy set for all countries. If a GATT member nation refuses to reduce its tariffs, then other members refuse to lower theirs. If a GATT member raises its tariffs on some product above the level to which

it had previously agreed, then the other member nations are allowed, under the agreement, to retaliate with increases in their own tariffs. In this way, nations have a greater incentive to move in the direction of free trade and a disincentive to take advantage of others by unilaterally raising their tariffs.

The simple prisoner's dilemma trade policy game therefore offers a simple explanation of the need for international organizations like the GATT or the World Trade Organization (WTO). These agreements may represent methods to achieve cooperative solutions between trading countries.

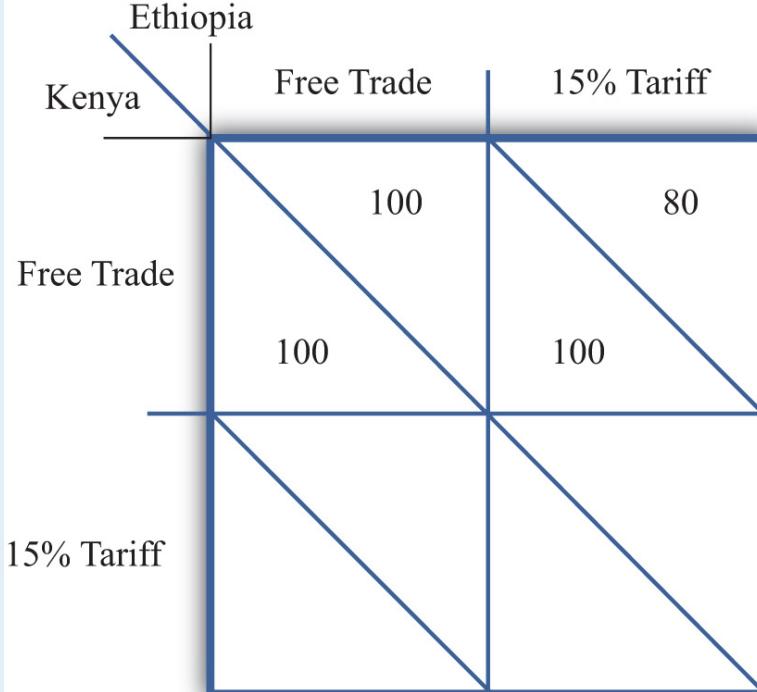
KEY TAKEAWAYS

- The goal of a noncooperative, or Nash, equilibrium in an optimal tariff game between two countries is for both countries to impose optimal tariffs.
- The goal of a cooperative equilibrium in an optimal tariff game between two countries is for both countries to set zero tariffs—that is, to choose free trade.
- The Nash equilibrium in an optimal tariff game between two countries is a “prisoner's dilemma” outcome because there is another set of strategies (not chosen) that could make both countries better off.
- The WTO, and the GATT before it, represents mechanisms by which countries can achieve the cooperative equilibrium.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a country’s countertrade actions in response to its trading partner’s increase in tariffs.
 - b. The name given to a noncooperative solution to an economic game.
 - c. The term used to describe an economic game equilibrium that maximizes the sum of the payoffs to all players.
2. Consider the following trade policy game between two small country governments, Kenya and Ethiopia. The policy choices for each government are either to choose free trade on all imports or to place a 15 percent tariff on all imports. The national welfare payoffs for each country when both choose free trade are given as (100, 100). The first 100 is Kenya’s national welfare; the second is Ethiopia’s.

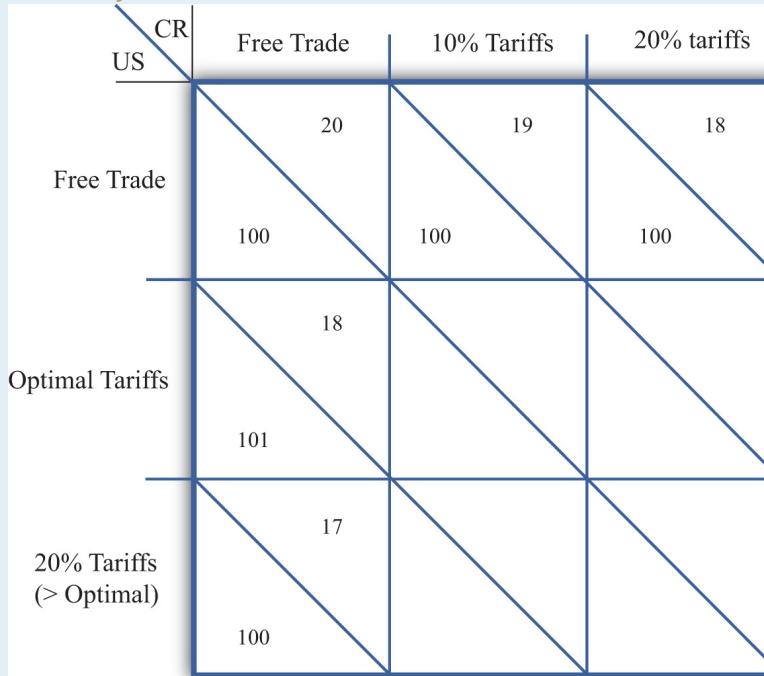
Figure 7.21
A Trade Policy Game



- a. Based on the tariff analysis for a small importing country and assuming symmetry between the two countries, complete the empty two cells in the table above.
 - b. Based on the numbers you provided in part a, identify which cell corresponds to the Nash (or noncooperative) equilibrium.
 - c. Which cell corresponds to the cooperative equilibrium?
 - d. Does this game help justify a trade liberalization organization like the WTO?
3. Suppose the United States (US) and Costa Rica (CR) are two countries among many others in the world. The US is a large country and thus its import tariffs will lower the price of CR's exports. CR, however, is a small country, so its tariffs do not affect prices in the US. Assume the US government can choose free trade, optimal tariffs, or 20 percent tariffs. CR can choose free trade, 10 percent tariffs, or 20 percent tariffs on all imports. The national welfare payoffs for each country in five cases are

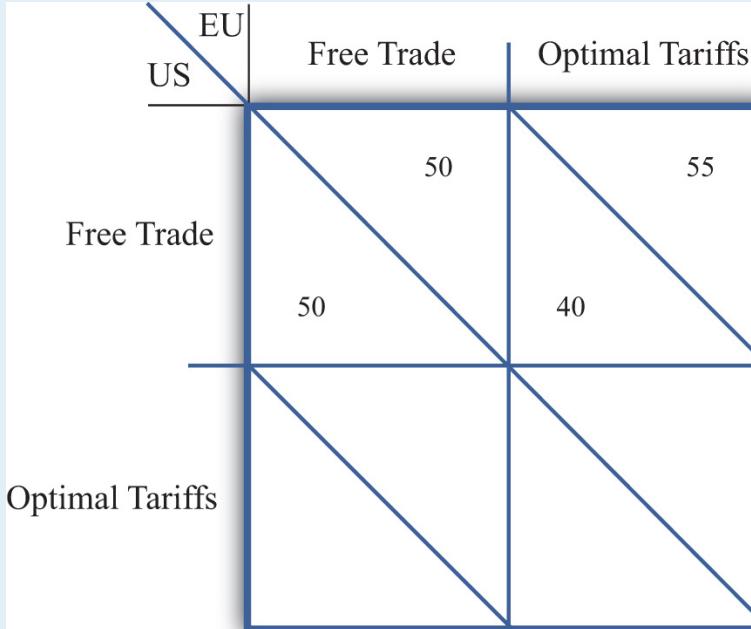
given. The first term is the US's national welfare; the second is CR's.

Figure 7.22
A Trade Policy Game



- Use the information provided in the table to complete the four empty cells.
 - Among the nine outcomes, which would CR most prefer?
 - Among the nine outcomes, which would the US most prefer?
 - Identify which cell or cells correspond to a Nash (or noncooperative) equilibrium.
 - Which cell corresponds to the cooperative equilibrium?
4. Consider the following trade policy game between two large country governments, the US and the EU. The policy choices for each government are to choose either free trade on all imports or to place an optimal tariff on all imports. The national welfare payoffs for each country when both choose free trade are given as (50, 50). The first term is the US's national welfare; the second is the EU's.

Figure 7.23
A Trade Policy Game



- Based on the tariff analysis for a large importing country and assuming symmetry between the two countries, complete the empty two cells in the table.
- Among the four outcomes, which would the US most prefer? Which would the EU most prefer?
- Identify which cell corresponds to the Nash (or noncooperative) equilibrium.
- Which cell corresponds to the cooperative equilibrium?
- Does this game help justify a trade liberalization organization like the WTO?

7.10 Import Quotas: Large Country Price Effects

LEARNING OBJECTIVES

1. Identify the effects of an import quota on prices in both countries and the quantity traded.
2. Know the equilibrium conditions that must prevail in a quota equilibrium.

Suppose Mexico, the importing country in free trade, imposes a binding import quota on wheat. The quota will restrict the flow of wheat across the border. As a result, the supply of wheat to the Mexican market will fall, and if the price remains the same, it will cause excess demand for wheat in the market. The excess demand will induce an increase in the price of wheat. Since wheat is homogeneous and the market is perfectly competitive, the price of all wheat sold in Mexico, both Mexican wheat and U.S. imports, will rise in price. The higher price will, in turn, reduce demand and increase domestic supply, causing a reduction in Mexico's import demand.

The restricted wheat supply to Mexico will shift supply back to the U.S. market. Since Mexico is assumed to be a large importer, the supply shifted back to the U.S. market will generate excess supply in the U.S. market at the original price and cause a reduction in the U.S. price. The lower price will, in turn, reduce U.S. supply, raise U.S. demand, and cause a reduction in U.S. export supply.

These price effects are identical in direction to the price effects of an import tax, a voluntary export restraint, and an export tax.

A new quota equilibrium will be reached when the following two conditions are satisfied:

$$MD^{Mex}(P_Q^{Mex}) = \bar{Q}$$

and

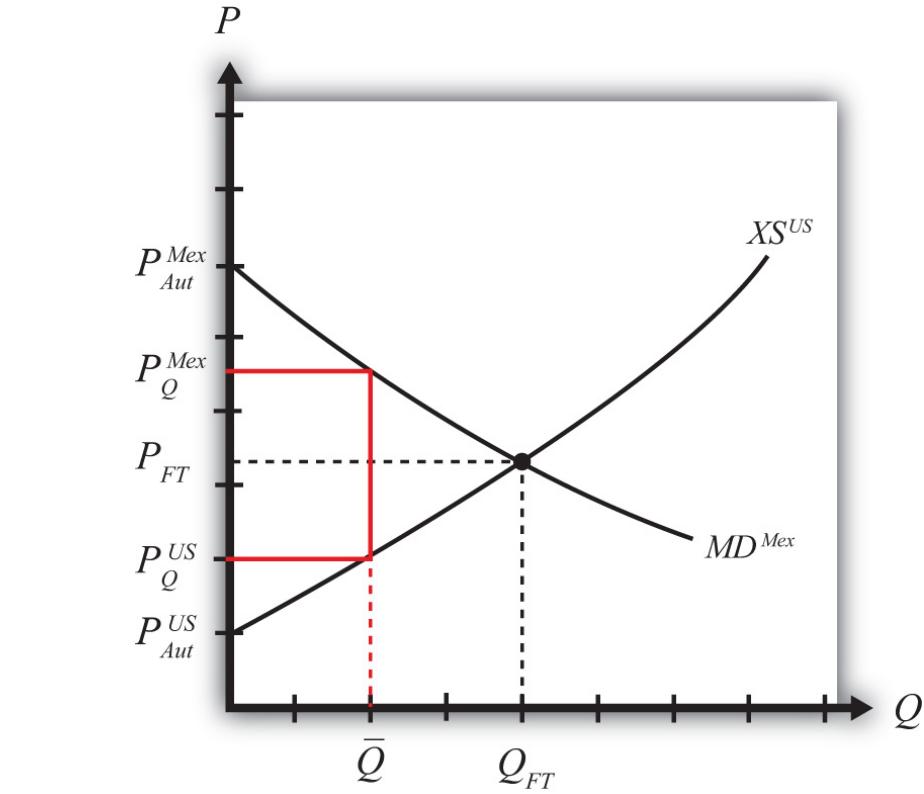
$$XS^{US}(P_Q^{US}) = \bar{Q},$$

where \bar{Q} is the quantity at which the quota is set, P_Q^{Mex} is the price in Mexico after the quota, and P_Q^{US} is the price in the United States after the quota.

The first condition says that the price must change in Mexico such that import demand falls to the quota level \bar{Q} . In order for this to occur, the price in Mexico rises. The second condition says that the price must change in the United States such that export supply falls to the quota level \bar{Q} . In order for this to occur, the price in the United States falls.

The quota equilibrium is depicted on the graph in [Figure 7.24 "Depicting a Quota Equilibrium: Large Country Case"](#). The Mexican price of wheat rises from P_{FT} to P_Q^{Mex} , which is sufficient to reduce its import demand from Q_{FT} to \bar{Q} . The U.S. price of wheat falls from P_{FT} to P_Q^{US} , which is sufficient to reduce its export supply from Q_{FT} to \bar{Q} .

Figure 7.24 Depicting a Quota Equilibrium: Large Country Case



Notice that there is a unique set of prices that satisfies the equilibrium conditions for every potential quota that is set. If the quota were set lower than \bar{Q} , the price

wedge would rise, causing a further increase in the Mexican price and a further decrease in the U.S. price.

At the extreme, if the quota were set equal to zero, then the prices in each country would revert to their autarky levels. In this case, the quota would *prohibit* trade.

KEY TAKEAWAYS

- An import quota will raise the domestic price and, in the case of a large country, lower the foreign price.
- The difference between the foreign and domestic prices after the quota is implemented is known as a quota rent.
- An import quota will reduce the quantity of imports to the quota amount.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The direction of change of domestic producer surplus when an import quota is implemented by a domestic country.
 - b. The direction of change of the domestic price after a binding import quota is implemented by a domestic country.
 - c. The direction of change of the foreign price after a binding import quota is implemented by a large domestic country.
 - d. Of *increase, decrease, or stay the same*, this is the effect on the domestic price after a nonbinding import quota is implemented by a domestic country.
 - e. The term used to describe a zero quota that eliminates trade.
 - f. Of *increase, decrease, or stay the same*, this is the effect on the price of U.S.-made automobiles if the United States restricts the quantity of imported foreign automobiles.
 - g. Of *increase, decrease, or stay the same*, this is the effect on the quantity of wheat imports if a binding import quota is implemented.
 - h. Of *increase, decrease, or stay the same*, this is the effect on foreign exports of wheat if a binding import quota is implemented by an importing country.

7.11 Administration of an Import Quota

LEARNING OBJECTIVE

1. Learn the different ways in which an import quota can be implemented to monitor and assure that only the specified amount is allowed to enter.

When a quantity restriction is set by a government, it must implement procedures to prevent imports beyond the restricted level. A binding import quota will result in a higher price in the import country and, in the case of a large country, a price reduction in the exporter's market. The price wedge would generate profit opportunities for anyone who could purchase (or produce) the product at the lower price (or cost) in the export market and resell it at the higher price in the import market.

Three basic methods are used to administer import quotas.

1. *Offer quota rights on a first-come, first-served basis.* The government could allow imports to enter freely from the start of the year until the quota is filled. Once filled, customs officials would prohibit entry of the product for the remainder of the year. If administered in this way, the quota may result in a fluctuating price for the product over the year. During the open period, a sufficient amount of imports may flow in to achieve free trade prices. Once the window is closed, prices would revert to the autarky prices.
2. *Auction quota rights.* Essentially, the government could sell quota tickets, where each ticket presented to a customs official would allow the entry of one unit of the good. If the tickets are auctioned, or if the price is determined competitively, the price at which each ticket would be sold is the difference in prices that exists between the export and import market. The holder of a quota ticket can buy the product at the low price in the exporter's market and resell it at the higher price in the importer's market. If there are no transportation costs, a quota holder can make a pure profit, called a quota rent, equal to the difference in prices. If the government sells the quota tickets at the maximum attainable price, then the government would receive all the quota rents.

3. *Give away quota rights.* The government could give away the quota rights by allocating quota tickets to appropriate individuals. The recipient of a quota ticket essentially receives a windfall profit since, in the absence of transportation costs, they can claim the entire quota rent at no cost to themselves. Governments often allocate the quota tickets to domestic importing companies based on past market shares. Thus, if an importer of the product had imported 20 percent of all imports prior to the quota, then it would be given 20 percent of the quota tickets. Sometimes governments give the quota tickets away to foreigners. In this case, the allocation acts as a form of foreign aid since the foreign recipients receive the quota rents. It is worth noting that because quota rents are so valuable, a government can use them to direct rents toward its political supporters.

KEY TAKEAWAYS

- To administer a quota, countries generally issue quota tickets, or import licenses, with the allowable import quantity limited in total to the quota level.
- The government earns revenue from the quota rents if it allocates the quota tickets via auction or sale.
- If the government gives the quota tickets away, the recipients of the quota tickets earn the quota rents.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *domestic* or *foreign* residents, this group receives quota rents when the government sells the right to import.
 - b. The term for the quota allocation method in which imports are allowed freely until the quota limit is reached.
 - c. The term used to describe the sale of quota rights to the highest bidder.
 - d. The likely recipients if new quota rights are given away by the government.
 - e. The term used to describe the profit made by a quota rights holder who can purchase the product cheaper in the export market and sell it for more in the import market.

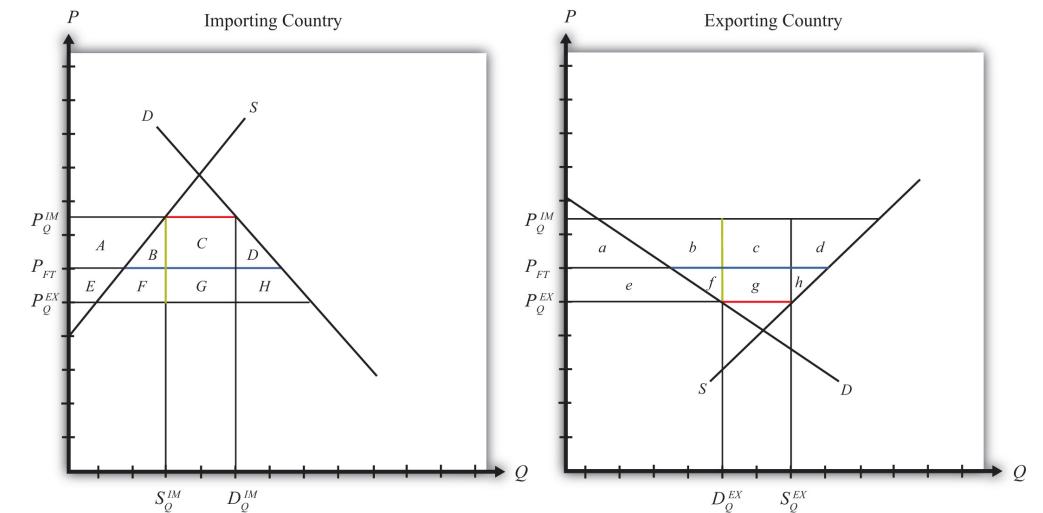
7.12 Import Quota: Large Country Welfare Effects

LEARNING OBJECTIVES

1. Use a partial equilibrium diagram to identify the welfare effects of an import quota on producer and consumer groups and the government in the importing and exporting countries.
2. Calculate the national and world welfare effects of an import quota.

Suppose for simplicity that there are only two trading countries: one importing country and one exporting country. The supply and demand curves for the two countries are shown in [Figure 7.25 "Welfare Effects of a Quota: Large Country Case"](#). P_{FT} is the free trade equilibrium price. At that price, the excess demand by the importing country equals the excess supply by the exporter.

Figure 7.25 Welfare Effects of a Quota: Large Country Case



The free trade quantity of imports and exports is shown as the blue line segment on each country's graph (the horizontal distance between the supply and demand curves at the free trade price). Suppose the large importing country implements a binding quota set equal to the length of the red line segment (the horizontal distance between the supply and demand curves at either the higher import price or the lower export price). When a new equilibrium is reached, the price in the importing country will rise until import demand is equal to the quota level. The

price in the exporting country will fall until export supply is equal to the quota level.

Table 7.5 "Welfare Effects of an Import Quota" provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the governments in the importing and exporting countries. The aggregate national welfare effects and the world welfare effects are also shown.

Table 7.5 Welfare Effects of an Import Quota

	Importing Country	Exporting Country
Consumer Surplus	$-(A + B + C + D)$	$+e$
Producer Surplus	$+A$	$-(e + f + g + h)$
Quota Rents	$+(C + G)$	0
National Welfare	$+G - (B + D)$	$-(f + g + h)$
World Welfare	$-(B + D) - (f + h)$	

Refer to Table 7.5 "Welfare Effects of an Import Quota" and Figure 7.25 "Welfare Effects of a Quota: Large Country Case" to see how the magnitude of the changes is represented.

Import quota effects on the importing country's consumers. Consumers of the product in the importing country suffer a reduction in well-being as a result of the quota. The increase in the domestic price of both imported goods and the domestic substitutes reduces the amount of consumer surplus in the market.

Import quota effects on the importing country's producers. Producers in the importing country experience an increase in well-being as a result of the quota. The increase in the price of their product on the domestic market increases producer surplus in the industry. The price increases also induce an increase in the output of existing firms (and perhaps the addition of new firms), an increase in employment, and an increase in profit, payments, or both to fixed costs.

Import quota effects on the quota rents. Who receives the quota rents depends on how the government administers the quota.

1. If the government auctions the quota rights for their full price, then the government receives the quota rents. In this case, the quota is

equivalent to a specific tariff set equal to the difference in prices ($T = P_Q^{IM} - P_Q^{EX}$), shown as the length of the green line segment in [Figure 7.25 "Welfare Effects of a Quota: Large Country Case"](#).

2. If the government gives away the quota rights, then the quota rents accrue to whoever receives these rights. Typically, they would be given to someone in the importing economy, which means that the benefits would remain in the domestic economy.
3. If the government gives the quota rights away to foreigners, then the foreigners receive the quota rents. This would imply that these rents should be shifted to the exporting country's effects and subtracted from the importing country's effects.

Import quota effects on the importing country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the recipients of the quota rents. Assume that the quota rent recipients are domestic residents. The net effect consists of three components: a positive terms of trade effect (G), a negative production distortion (B), and a negative consumption distortion (D).

Because there are both positive and negative elements, the net national welfare effect can be either positive or negative. The interesting result, however, is that it can be *positive*. This means that a quota implemented by a large importing country *may raise* national welfare.

Generally speaking, the following are true:

1. Whenever a large country implements a small restriction on imports, it will raise national welfare.
2. If the quota is too restrictive, national welfare will fall.
3. There will be a positive quota level that will maximize national welfare.

However, it is also important to note that not everyone's welfare rises when there is an increase in national welfare. Instead, there is a redistribution of income. Producers of the product and recipients of the quota rents will benefit, but consumers will lose. A national welfare increase, then, means that the sum of the gains exceeds the sum of the losses across all individuals in the economy. Economists generally argue that, in this case, compensation from winners to losers can potentially alleviate the redistribution problem.

Import quota effects on the exporting country's consumers. Consumers of the product in the exporting country experience an increase in well-being as a result of the quota.

The decrease in their domestic price raises the amount of consumer surplus in the market.

Import quota effects on the exporting country's producers. Producers in the exporting country experience a decrease in well-being as a result of the quota. The decrease in the price of their product in their own market decreases producer surplus in the industry. The price decline also induces a decrease in output, a decrease in employment, and a decrease in profit, payments, or both to fixed costs.

Import quota effects on the quota rents. There are no quota rent effects on the exporting country as a result of the importer's quota unless the importing government gives away the quota rights to foreigners. Only in this case would the rents accrue to someone in the exporting country.

Import quota effects on the exporting country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers and producers. The net effect consists of three components: a negative terms of trade effect (g), a negative consumption distortion (f), and a negative production distortion (h).

Since all three components are negative, the importer's tariff must result in a reduction in national welfare for the exporting country. However, it is important to note that a redistribution of income occurs—that is, some groups gain while others lose. In this case, the sum of the losses exceeds the sum of the gains.

Import quota effects on world welfare. The effect on world welfare is found by summing the national welfare effects on the importing and exporting countries. By noting that the terms of trade gain to the importer is equal to the terms of trade loss to the exporter, the world welfare effect reduces to four components: the importer's negative production distortion (B), the importer's negative consumption distortion (D), the exporter's negative consumption distortion (f), and the exporter's negative production distortion (h). Since each of these is negative, the world welfare effect of the import quota is negative. The sum of the losses in the world exceeds the sum of the gains. In other words, we can say that an import quota results in a reduction in world production and consumption efficiency.

KEY TAKEAWAYS

- An import quota lowers consumer surplus in the import market and raises it in the export country market.
- An import quota raises producer surplus in the import market and lowers it in the export country market.
- National welfare may rise or fall when a large country implements an import quota.
- National welfare in the exporting country falls when an importing country implements an import quota.
- An import quota of any size will reduce world production and consumption efficiency and thus cause world welfare to fall.

EXERCISES

1. Consider the following trade policy action (applied by the domestic country) listed at the top of the second column in the table below. In the empty boxes, use the following notation to indicate the effect of the policy on the variables listed in the first column:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

Use a partial equilibrium model to determine the answers, and assume that the shapes of the supply and demand curves are “normal.” Assume that the policy does not begin with, or result in, prohibitive trade policies. Also assume that the policy does not correct for market imperfections or distortions.

For example, an import quota applied by a large country will cause an increase in the domestic price of the import good; therefore a + is placed in the first box of the table.

TABLE 7.6 IMPORT QUOTA EFFECTS

	An Import Quota by a Large Country Initially in Free Trade
Domestic Market Price	+
Domestic Industry Employment	

	An Import Quota by a Large Country Initially in Free Trade
Domestic Consumer Welfare	
Domestic Producer Welfare	
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	
Foreign Consumer Welfare	
Foreign Producer Welfare	
Foreign National Welfare	

2. Suppose there are two large countries, the United States and China. Assume that both countries produce and consume clothing. The United States imports clothing from China. Consider the trade policy action listed at the top of the second column in the table below. In the boxes, indicate the effect of the policy on the variables listed in the first column. Use a partial equilibrium, perfect competition model to determine the answers. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 7.7 IMPORT QUOTA ELIMINATION
EFFECTS

	I
	Elimination of a U.S. Import Quota on Clothing Imports
U.S. Domestic Consumer Welfare	
U.S. Domestic Producer Welfare	
U.S. National Welfare	
Chinese Producer Welfare	
Chinese Consumer Welfare	
Chinese National Welfare	

7.13 Import Quota: Small Country Price Effects

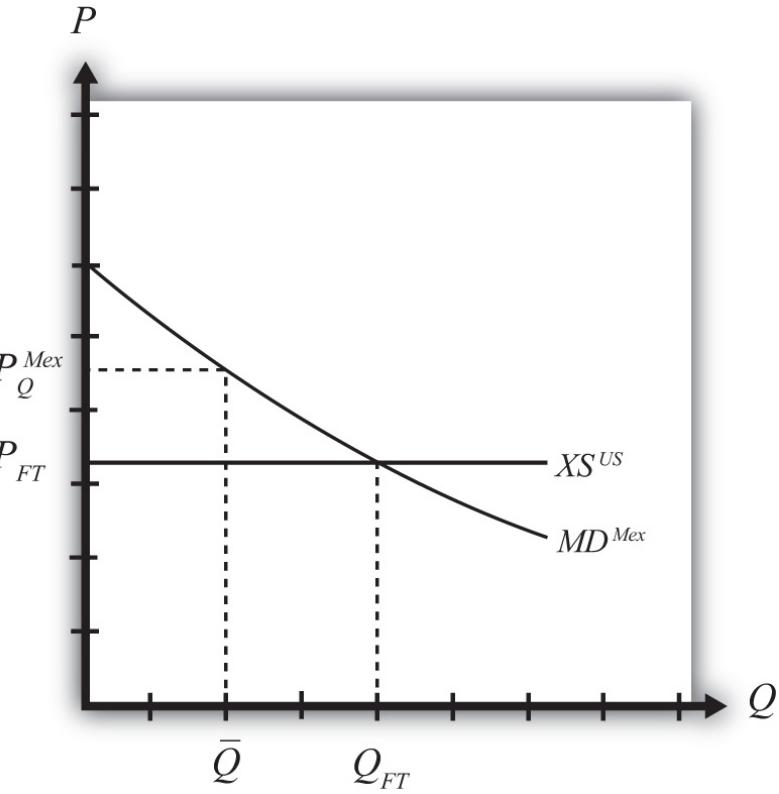
LEARNING OBJECTIVES

1. Identify the effects of an import quota on prices in both countries and the quantity traded in the case of a small country.
2. Know the equilibrium conditions that must prevail in a quota equilibrium.

The small country assumption means that the country's imports are a very small share of the world market—so small that even a complete elimination of imports would have an imperceptible effect on world demand for the product and thus would not affect the world price. Thus when a quota is implemented by a small country, there is no effect on the world price.

To depict the price effects of a quota, we use an export supply/import demand diagram shown in [Figure 7.26 "Depicting a Quota Equilibrium: Small Country Case"](#). The export supply curve is drawn as a horizontal line since the exporting country is willing to supply as much as the importer demands at the world price. The small importing country takes the world price as exogenous since it can have no effect on it.

Figure 7.26 Depicting a Quota Equilibrium: Small Country Case



When the quota is placed on imports, it restricts supply to the domestic market since fewer imports are allowed in. The reduced supply raises the domestic price. The world price is unaffected by the quota and remains at the free trade level. In the final equilibrium, two conditions must hold—the same two conditions as in the case of a large country, namely,

$$MD^{Mex}(P_Q^{Mex}) = \bar{Q}$$

and

$$XS^{US}(P_{FT}) = \bar{Q}.$$

This implies that, in the case of a small country, the price of the import good in the importing country must rise to the level at which the import demand is equal to the quota level. Export supply merely falls to the lower level now demanded.

KEY TAKEAWAYS

- An import quota will raise the domestic price and, in the case of a small country, leave the foreign price unchanged.
- A binding import quota will reduce the quantity of imports.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the effect on the world price when a small country implements a binding import quota.
 - b. Of *increase, decrease, or stay the same*, the effect on the import volume of a product when a small country implements a binding import quota.
 - c. Of *increase, decrease, or stay the same*, the effect on the exports from the rest of the world when a small country implements a binding import quota.

7.14 Import Quota: Small Country Welfare Effects

LEARNING OBJECTIVES

1. Use a partial equilibrium diagram to identify the welfare effects of an import quota on producer and consumer groups and the government in the importing country.
2. Calculate the national welfare effects of an import quota.

Consider a market in a small importing country that faces an international or world price of P_{FT} in free trade. The free trade equilibrium is depicted in [Figure 7.27 "Welfare Effects of a Quota: Small Country Case"](#), where P_{FT} is the free trade equilibrium price. At that price, domestic demand is given by D_{FT} , domestic supply by S_{FT} , and imports by the difference, $D_{FT} - S_{FT}$ (the blue line in the figure).

Suppose an import quota is set below the free trade level of imports. A reduction in imports will lower the supply on the domestic market and raise the domestic price. In the new equilibrium, the domestic price will rise to the level at which import demand equals the value of the quota. Since the country is small, there will be no effect on the world price, which will remain at P_{FT} .

In [Figure 7.27 "Welfare Effects of a Quota: Small Country Case"](#), if the quota is set equal to $\bar{Q} = D_Q - S_Q$ (the red line segment), then the price will have to rise to P_Q .

Figure 7.27 Welfare Effects of a Quota: Small Country Case

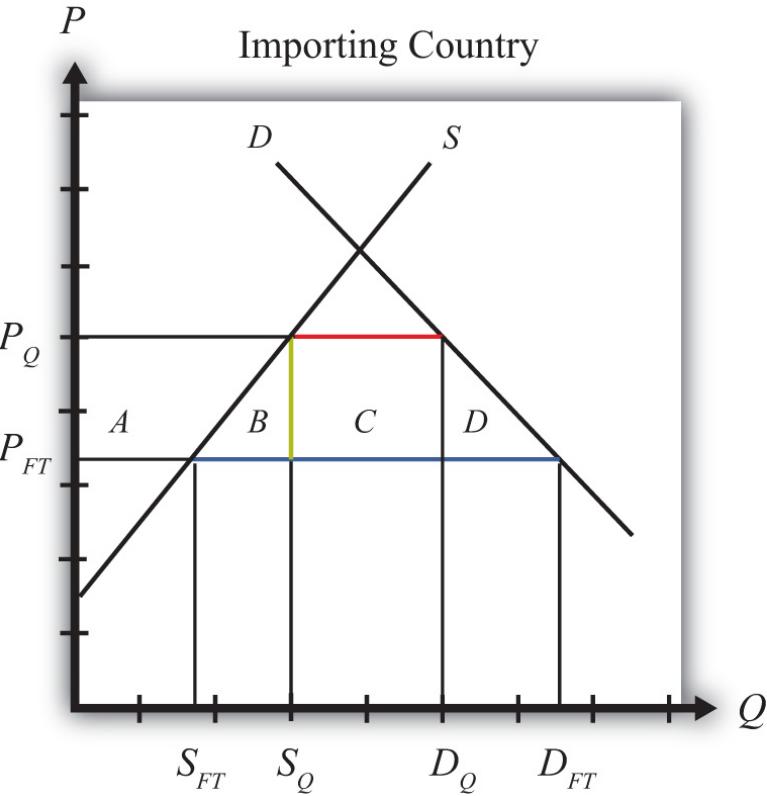


Table 7.8 "Welfare Effects of an Import Tariff" provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the recipients of the quota rents in the importing country. The aggregate national welfare effects are also shown.

Table 7.8 Welfare Effects of an Import Tariff

	Importing Country
Consumer Surplus	- (A + B + C + D)
Producer Surplus	+ A
Quota Rents	+ C
National Welfare	- B - D

Refer to [Table 7.8 "Welfare Effects of an Import Tariff"](#) and [Figure 7.27 "Welfare Effects of a Quota: Small Country Case"](#) to see how the magnitudes of the changes are represented.

Welfare effects on the importing country's consumers. Consumers of the product in the importing country are worse off as a result of the quota. The increase in the domestic price of both imported goods and the domestic substitutes reduces consumer surplus in the market.

Welfare effects on the importing country's producers. Producers in the importing country are better off as a result of the quota. The increase in the price of their product increases producer surplus in the industry. The price increase also induces an increase in the output of existing firms (and perhaps the addition of new firms), an increase in employment, and an increase in profit, payments, or both to fixed costs.

Welfare effects on the quota rents. Who receives the quota rents depends on how the government administers the quota.

1. If the government auctions the quota rights for their full price, then the government receives the quota rents. In this case, the quota is equivalent to a specific tariff set equal to the difference in prices ($t = P_Q - P_{FT}$), shown as the length of the green line segment in [Figure 7.27 "Welfare Effects of a Quota: Small Country Case"](#).
2. If the government gives away the quota rights, then the quota rents accrue to whoever receives these rights. Typically, they would be given to someone in the importing economy, which means that the benefits would remain in the domestic economy.
3. If the government gives the quota rights away to foreigners, then people in the foreign country receive the quota rents. In this case, the rents would not be a part of the importing country effects.

Welfare effects on the importing country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the domestic recipients of the quota rents. The net effect consists of two components: a negative production efficiency loss (B) and a negative consumption efficiency loss (D). The two losses together are referred to as “deadweight losses.”

Because there are only negative elements in the national welfare change, the net national welfare effect of a quota must be negative. This means that a quota implemented by a small importing country *must* reduce national welfare.

Generally speaking, the following are true:

1. Whenever a small country implements a quota, national welfare falls.
2. The more restrictive the quota, the larger will be the loss in national welfare.
3. The quota causes a redistribution of income. Producers and the recipients of the quota rents gain, while consumers lose.
4. Because the country is assumed to be small, the quota has no effect on the price in the rest of the world; therefore there are no welfare changes for producers or consumers there. Even though imports are reduced, the related reduction in exports by the rest of the world is assumed to be too small to have a noticeable impact.

KEY TAKEAWAYS

- An import quota lowers consumer surplus in the import market.
- An import quota by a small country has no effect on the foreign country.
- The national welfare effect of an import tariff is evaluated as the sum of the producer and consumer surplus and government revenue effects.
- An import quota of any size will result in deadweight losses and reduce production and consumption efficiency.
- National welfare falls when a small country implements an import quota.

EXERCISE

1. Consider the following trade policy action (applied by the domestic country) listed along the top row of the table below. In the boxes, indicate the effect of the policy on the variables listed in the first column. Use a partial equilibrium model to determine the answers. You do not need to show your work. Assume that the policy does not begin with, or result in, prohibitive trade policies. Also assume that the policy does not correct for market imperfections or distortions. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 7.9 IMPORT QUOTA EFFECTS

	Import Quota (Administered by Giving Away Quota Tickets) by a Small Country
Domestic Price	
Domestic Consumer Welfare	
Domestic Producer Welfare	
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	

	Import Quota (Administered by Giving Away Quota Tickets) by a Small Country
Foreign Consumer Welfare	
Foreign Producer Welfare	
Foreign National Welfare	

7.15 The Choice between Import Tariffs and Quotas

LEARNING OBJECTIVES

1. Understand the pros and cons of applying tariffs versus quotas.
2. Learn how tariffs differ from quotas in their protective effects in the face of market changes.

There are two basic ways to provide protection to domestic import-competing industries: a tariff or a quota. The choice between one or the other is likely to depend on several concerns.

One concern is the revenue effects. A tariff has an immediate advantage for governments in that it will automatically generate tariff revenue (assuming the tariff is not prohibitive). Quotas may or may not generate revenue depending on how the quota is administered. If a quota is administered by selling quota tickets (i.e., import rights), then a quota will generate government revenue; however, if the quota is administered on a first-come, first-served basis or if quota tickets are given away, then no revenue is collected.

Administrative costs of tariffs and quotas are also likely to differ. Tariff collection involves product identification, collection, and processing of fees. Quota administration will also involve product identification and some method of keeping track of, or counting, the product as it enters the country in multiple ports of entry. It may also involve some method of auctioning or disbursing quota tickets. It is not obvious which of these two procedures would be less costly, although a good guess would be tariff collection.

Perhaps the most important distinction between the two policies, however, is the protective effect the policy has on the import-competing industries. In one sense, quotas are more protective of the domestic industry because they limit the extent of import competition to a fixed maximum quantity. The quota provides an upper bound to the foreign competition the domestic industries will face. In contrast, tariffs simply raise the price but do not limit the degree of competition or trade volume to any particular level.

In the original General Agreement on Tariffs and Trade (GATT), a preference for the application of tariffs rather than quotas was introduced as a guiding principle. One

reason was the sense that tariffs allowed for more market flexibility and thus could be expected to be less protective over time. Another reason concerned transparency. With a quota in place, it is very difficult to discern the degree to which a market is protected since it can be difficult to measure how far the quota is below the free trade import level. With a tariff in place, especially an ad valorem tariff, one can use the tariff percentage as a measure of the degree of protection.

Also, it was considered somewhat easier to negotiate reductions in tariff rates than quota increases during GATT rounds of trade liberalization. Again, the issue of transparency arises. Trade liberalization agreements generally target a fixed percentage for tariff reductions. For example, countries might agree to reduce average tariffs by 30 percent from their current levels. This rule would be perceived as being equal reciprocation in that each country would be liberalizing to the same degree. Hence the agreement could be judged to be fair. However, with quotas in place, it would be difficult, if not impossible, to apply such a straightforward type of fairness principle.

For this reason, current World Trade Organization (WTO) member countries agreed in the Uruguay Round to phase out the use of quotas, used primarily in agriculture industries. Instead, countries will apply tariffs that are equivalent in their market effects to the original quotas. This adjustment is referred to as tariffication. In this way, future rounds of trade liberalization negotiations will be able to use fair reciprocal concessions to bring these tariffs down further.

The Protective Effects of Tariffs versus Quotas with Market Changes

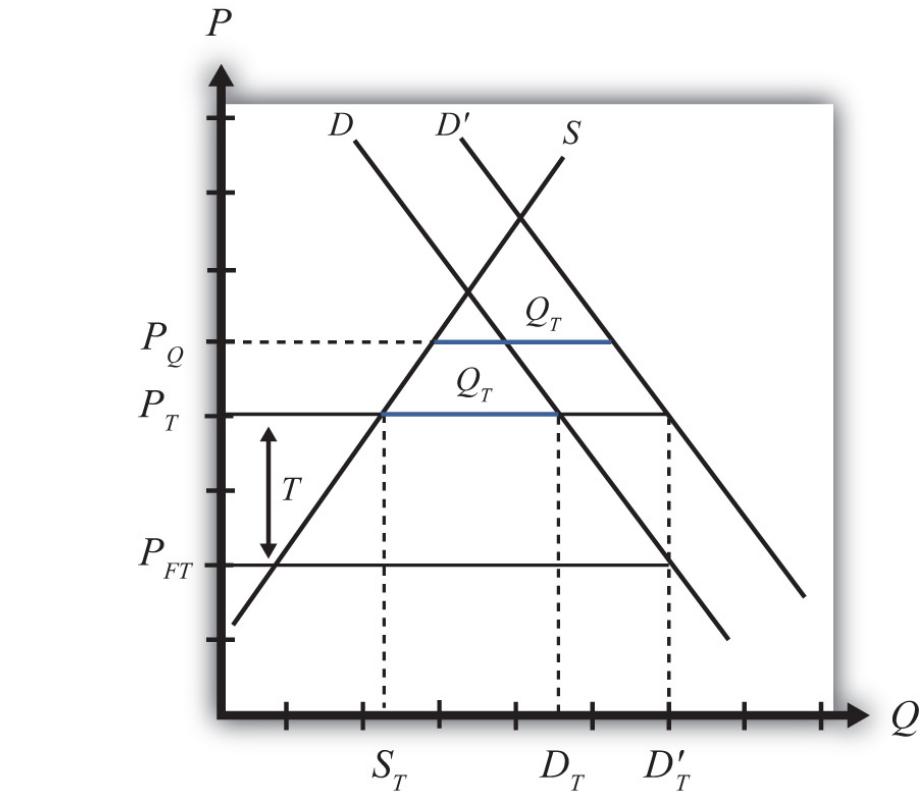
One of the main concerns in choosing between tariffs or quotas is the protective effect of the policy. Although tariffs and quotas are generally equivalent to each other in terms of their static price and welfare effects, this equivalence does not remain true in the face of market changes. In the next sections we consider three such market changes: an increase in domestic demand, an increase in domestic supply, and a decrease in the world price. In each case, we compare the protective effects of a tariff and a quota for the domestic import-competing industries.

An Increase in Domestic Demand

Consider [Figure 7.28 "Effects of a Demand Increase"](#), which depicts a small importing country. P_{FT} is the free trade price. If a tariff of T is put into place, the domestic price rises to P_T and imports equal $D_T - S_T$. A quota set equal to Q_T (the blue line segment) would generate the same increase in price to P_T and the same

level of imports. Thus the tariff T and quota Q_T are said to be equivalent to each other.

Figure 7.28 Effects of a Demand Increase



Next, consider the effects in this market when there is an increase in domestic demand, represented by a rightward shift of the demand curve. A demand increase could arise because of rising incomes in the country or because consumers' preferences become more favorable to this product.

With a tariff in place initially, the increase in domestic demand will leave the domestic price unaffected. Because this is a small country, the world price does not change and thus the domestic tariff-inclusive price remains at $P_T = P_{FT} + T$. Domestic supply also remains at S_T , but demand rises to D'_T , causing an increase in imports to $D'_T - S_T$.

With a quota in place initially, the increase in domestic demand causes the domestic price to rise to P_Q in order to maintain the import level at Q_T (the higher blue line

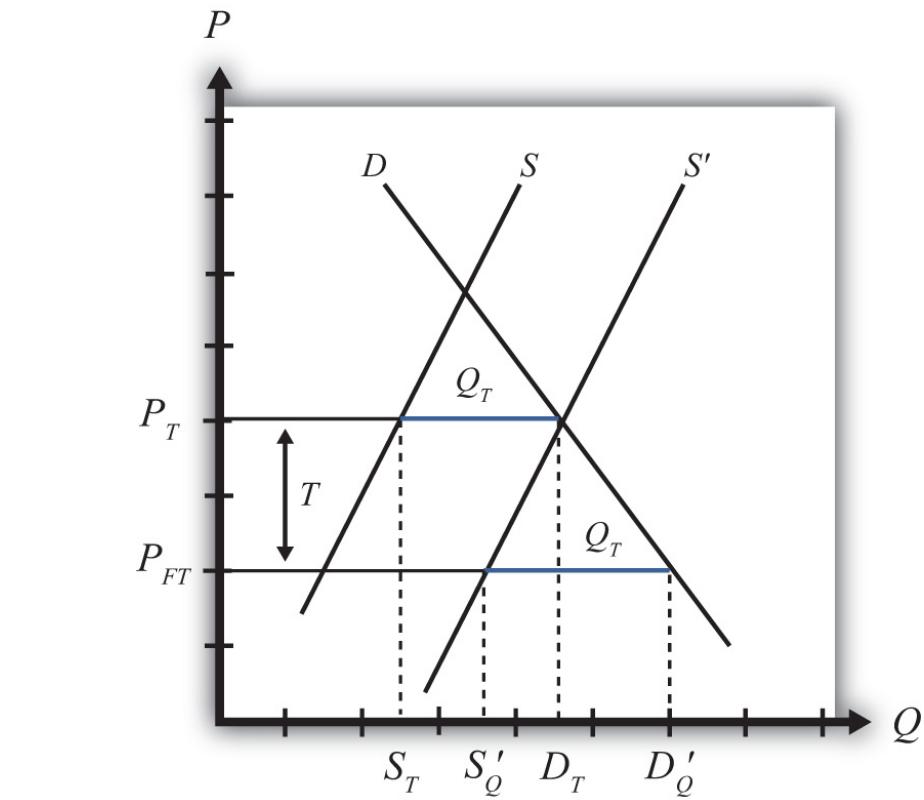
segment). Domestic supply will rise with the increase in price (not labeled), while domestic demand will fall.

The protective effect of the tariff or quota means the degree to which the domestic *producers* are protected in the face of the market change. Since the domestic price rises more with the quota in place than with the tariff, domestic producers will enjoy a larger supply and consequently a higher level of producer surplus (not shown). Thus the quota is more protective than a tariff in the face of an increase in domestic demand.

An Increase in Domestic Supply

Again, consider a small importing country. In [Figure 7.29 "Effects of a Supply Increase"](#), P_{FT} is the free trade price. If a tariff of T is put into place, the domestic price rises to P_T and imports equal $D_T - S_T$. A quota set equal to Q_T (the blue line segment) would generate the same increase in price to P_T and the same level of imports. Thus the tariff T and quota Q_T are said to be equivalent to each other.

Figure 7.29 Effects of a Supply Increase



Next, consider the effects in this market when there is an increase in domestic supply, represented by a rightward shift of the supply curve. A supply increase could arise because of falling production costs or due to improvements in productivity.

With a tariff in place initially, the increase in domestic supply will leave the domestic price unaffected. Because this is a small country, the world price does not change and thus the domestic tariff-inclusive price remains at $P_T = P_{FT} + T$. However, because domestic supply is now higher at every price, at the price P_T , supply equals domestic demand of D_T . This means that with the tariff, imports are reduced to zero.

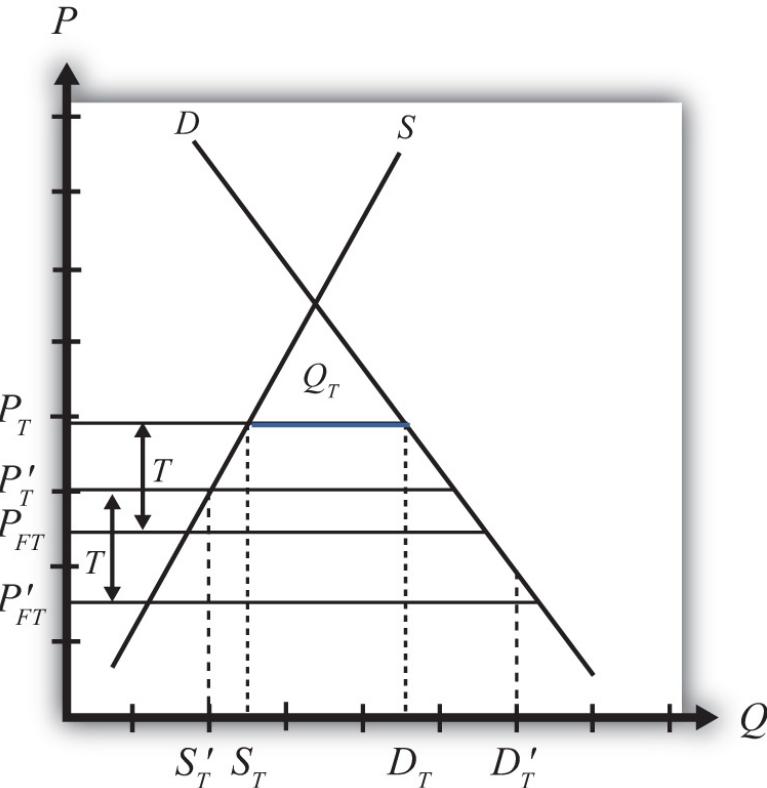
With a quota in place initially, the increase in domestic supply causes the domestic price to fall back to the free trade level in order to maintain the import level at the level Q_T (the lower blue line segment). Domestic supply will rise to S'_Q with the decrease in price, while domestic demand also will rise to D'_Q .

Since the domestic price rises more with the tariff in place than with the quota, domestic producers will enjoy a larger supply (D_T vs. S'_Q) and consequently a higher level of producer surplus (not shown). Thus the tariff is more protective than a quota in the face of an increase in domestic supply.

A Decrease in the World Price

Again, consider a small importing country. In [Figure 7.30 "Effects of a World Price Decrease"](#), P_{FT} is the free trade price. If a tariff of T is put into place, the domestic price rises to P_T and imports equal $D_T - S_T$. A quota set equal to Q_T (the blue line segment) would generate the same increase in price to P_T and the same level of imports. Thus the tariff T and quota Q_T are said to be equivalent to each other.

Figure 7.30 Effects of a World Price Decrease



Next, consider the effects in this market when there is a decrease in the world free trade price, represented by a downward shift from P_{FT} to P'_{FT} . The world price could fall because of falling world production costs or due to improvements in foreign productivity.

With a tariff in place initially, the decrease in the world price will cause a reduction in the domestic price. Because this is a small country, when the world price falls, the domestic tariff-inclusive price also falls to $P'_T = P'_{FT} + T$. With the lower price, domestic supply falls to S'_T , while domestic demand rises to D'_T . This means that with the tariff in place, imports rise to $D'_T - S'_T$.

With a quota in place initially, the decrease in the world free trade price has no effect on the domestic price. The domestic price remains at P_T since this is the only price that will support the quota Q_T .

Since the domestic price is higher with the quota in place than with the tariff, domestic producers will enjoy a larger supply (S_T vs. S'_T) and consequently a higher

level of producer surplus (not shown). Thus the quota is more protective than a tariff in the face of a decrease in the world free trade price.

The General Rule

What we can conclude from the three examples above is that when market conditions change such that imports increase, a quota is more protective than a tariff. This will occur if domestic demand increases, domestic supply decreases, the world price falls, or if some combination of these things occur.

In situations where market changes cause a decrease in imports, a tariff is more protective than a quota. This occurs if domestic demand falls, domestic supply rises, the world price rises, or some combination of these changes occurs.

Since protection is often provided due to the insistence of the domestic import-competing industries—rather than a more comprehensive concern for the general welfare of the country—and since import-competing firms are generally more concerned about situations where imports may increase, industry preferences usually favor quotas over tariffs since quotas will be more protective in these situations. Other government concerns, such as revenue needs, ease of administration, or participation in trade agreements like the GATT/WTO, which contain a preference of tariffs over quotas, have resulted in the widespread application of tariffs rather than quotas in most instances.

KEY TAKEAWAYS

- The effects of tariffs are more transparent than quotas and hence are a preferred form of protection in the GATT/WTO agreement.
- A quota is more protective of the domestic import-competing industry in the face of import volume increases.
- A tariff is more protective in the face of import volume decreases.

EXERCISES

1. Draw a diagram depicting a small importing country with a nonprohibitive import tariff (T) in place. On the diagram indicate the tariff rate and the equivalent import quota (Q) that would generate the same domestic price.

Next, suppose there is a decrease in domestic demand for the good.

- a. Indicate on the graph the new equilibrium with the tariff in place and the quota in place.
 - b. Indicate the new level of imports with the tariff and the quota. Which is larger?
 - c. Indicate the new domestic price with the tariff and the quota. Which is higher?
 - d. Which is more protective of the domestic import-competing industry in this situation, a tariff or quota? Explain why.
2. Draw a diagram depicting a small importing country with a nonprohibitive import tariff (T) in place. On the diagram indicate the tariff rate and the equivalent import quota (Q) that would generate the same domestic price.

Next, suppose there is an increase in the world price of the good.

- a. Indicate on the graph the new equilibrium with the tariff in place and the quota in place.
- b. Indicate the new level of imports with the tariff and the quota. Which is larger?
- c. Indicate the new domestic price with the tariff and the quota. Which is higher?
- d. Which is more protective of the domestic import-competing industry in this situation, a tariff or quota? Explain why.

7.16 Export Subsidies: Large Country Price Effects

LEARNING OBJECTIVES

1. Identify the effects of an export subsidy on prices in both countries and the quantity traded in a large country.
2. Know the equilibrium conditions that must prevail in a subsidy equilibrium.

Suppose the United States, the exporting country in free trade, implements a specific export subsidy on exports of wheat. A subsidy to exports will encourage the flow of wheat across the border. It will now cost less to move the product from the United States into Mexico.

As a result, the supply of wheat to the Mexican market will rise, causing a decrease in the price of wheat. Since the United States is assumed to be a large country, the price of all wheat sold in Mexico, both Mexican wheat and U.S. imports, will fall in price. The lower price will raise Mexico's import demand.

The higher wheat supply to Mexico will reduce supply in the U.S. market and induce an increase in the U.S. price. The higher price will raise U.S. export supply.

A new subsidy-ridden equilibrium will be reached when the following two conditions are satisfied:

$$P_S^{US} = P_S^{Mex} + S$$

and

$$XS^{US}(P_S^{US}) = MD^{Mex}(P_S^{Mex}),$$

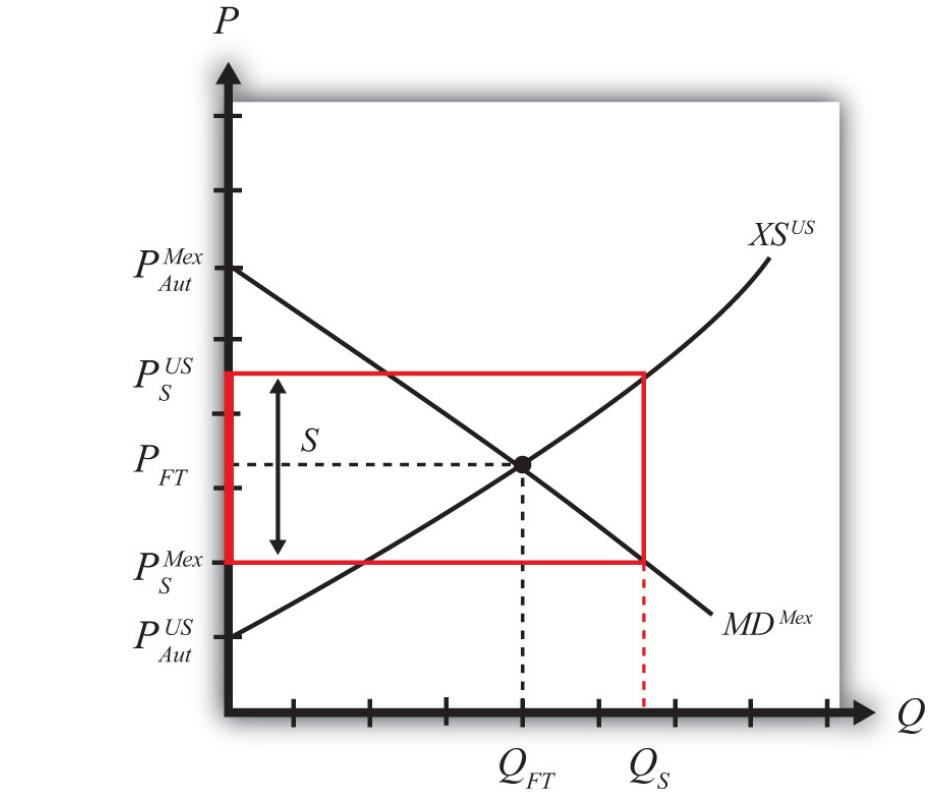
where S is the specific export subsidy, P_S^{Mex} is the price in Mexico after the subsidy, and P_S^{US} is the price in the United States after the subsidy. The first condition represents a price wedge between the final U.S. price and the Mexican price equal to the amount of the export subsidy. The prices must differ by the subsidy because U.S. suppliers of wheat must receive the same price for their product, regardless of whether the product is sold in the United States or Mexico, and all wheat sold in

Mexico must be sold at the same price. Since a subsidy is paid to U.S. exporters, the only way for these price equalities within countries to arise is if the price differs across countries by the amount of the subsidy.

The second condition states that the amount the United States wants to export at its new higher price must be equal to the amount Mexico wants to import at its new lower price. This condition guarantees that world supply of wheat equals world demand for wheat.

The export subsidy equilibrium is depicted graphically in [Figure 7.31 "Depicting a Subsidy Equilibrium: Large Country Case"](#). The Mexican price of wheat falls from P_{FT} to P_S^{Mex} , which raises its import demand from Q_{FT} to Q_S . The Mexican price of wheat falls from P_{FT} to P_S^{Mex} , which raises its import demand from Q_{FT} to Q_S . The U.S. price of wheat rises from P_{FT} to P_S^{US} , which raises its export supply also from Q_{FT} to Q_S . The difference in the prices between the two markets is equal to the export subsidy rate S .

Figure 7.31 Depicting a Subsidy Equilibrium: Large Country Case



KEY TAKEAWAYS

- An export subsidy will raise the domestic price and, in the case of a large country, reduce the foreign price.
- An export subsidy will increase the quantity of exports.
- The export subsidy will drive a price wedge, equal to the subsidy value, between the foreign price and the domestic price of the product.
- With the export subsidy in place in a two-country model, export supply at the higher domestic price will equal import demand at the lower foreign price.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The direction of change of the foreign price of soybeans when a large domestic country implements an export subsidy.
 - b. The direction of change of the domestic price of corn when a large domestic country implements an export subsidy.
 - c. The price of tea in the exporting country if the large exporter sets a subsidy of \$0.45 per pound and if the importer country price is \$3.25 inclusive of the subsidy abroad.
 - d. Of *increase, decrease, or stay the same*, this is the effect on the quantity of wheat produced domestically when an export subsidy is implemented by a large exporter.
 - e. Of *increase, decrease, or stay the same*, this is the effect on imports of wheat abroad if a wheat subsidy is implemented by a large exporting country.
 - f. Of *increase, decrease, or stay the same*, this is the effect on domestic consumption of cotton if a cotton export subsidy is implemented by a large exporting country.

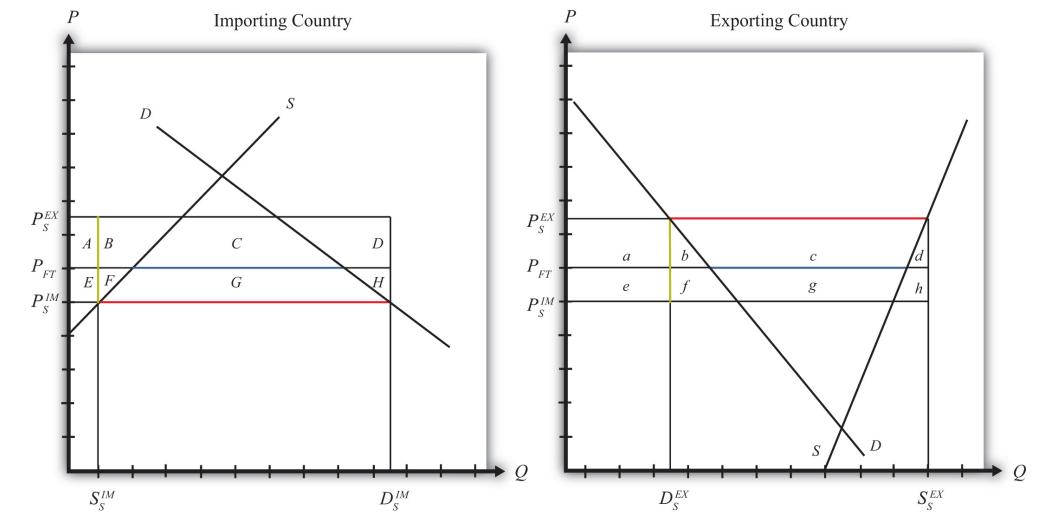
7.17 Export Subsidies: Large Country Welfare Effects

LEARNING OBJECTIVES

1. Use a partial equilibrium diagram to identify the welfare effects of an export subsidy on producer and consumer groups and the government in the exporting and importing countries.
2. Calculate the national and world welfare effects of an export subsidy.

Suppose that there are only two trading countries: one importing country and one exporting country. The supply and demand curves for the two countries are shown in [Figure 7.32 "Welfare Effects of a Subsidy: Large Country Case"](#). P_{FT} is the free trade equilibrium price. At that price, the excess demand by the importing country equals the excess supply by the exporter.

Figure 7.32 Welfare Effects of a Subsidy: Large Country Case



The quantity of imports and exports is shown as the blue line segment on each country's graph (the horizontal distance between the supply and demand curves at the free trade price). When a large exporting country implements an export subsidy, it will cause an increase in the price of the good on the domestic market and a decrease in the price in the rest of the world (RoW). Suppose after the subsidy the price in the importing country falls to P_T^{IM} and the price in the exporting country rises to P_T^{EX} . If the subsidy is a specific subsidy, then the subsidy rate would

be $S = P_S^{EX} - P_S^{IM}$, equal to the length of the green line segment in [Figure 7.32 "Welfare Effects of a Subsidy: Large Country Case"](#).

[Table 7.10 "Welfare Effects of an Export Subsidy"](#) provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the governments in the importing and exporting countries. The aggregate national welfare effects and the world welfare effects are also shown.

Table 7.10 Welfare Effects of an Export Subsidy

	Importing Country	Exporting Country
Consumer Surplus	+ (E + F + G)	- (a + b)
Producer Surplus	- (E + F)	+ (a + b + c)
Govt. Revenue	0	- (b + c + d + f + g + h)
National Welfare	+ G	- (b + d + f + g + h)
World Welfare	- (F + H) - (b + d)	

Refer to [Table 7.10 "Welfare Effects of an Export Subsidy"](#) and [Figure 7.32 "Welfare Effects of a Subsidy: Large Country Case"](#) to see how the magnitudes of the changes are represented.

Export subsidy effects on the exporting country's consumers. Consumers of the product in the exporting country experience a decrease in well-being as a result of the export subsidy. The increase in their domestic price lowers the amount of consumer surplus in the market.

Export subsidy effects on the exporting country's producers. Producers in the exporting country experience an increase in well-being as a result of the subsidy. The increase in the price of their product in their own market raises producer surplus in the industry. The price increase also induces an increase in output, an increase in employment, and an increase in profit, payments, or both to fixed costs.

Export subsidy effects on the exporting country's government. The government must pay the subsidy to exporters. These payments must come out of the general government budget. Who loses as a result of the subsidy payments depends on how the revenue is collected. If there is no change in total spending when the subsidy payments are made, then a reallocation of funds implies that funding to some other government program is reduced. If the subsidy is funded by raising tax revenues, then the

individuals responsible for the higher taxes lose out. If the government borrows money to finance the subsidy payments, then the budget cut or the tax increase can be postponed until some future date. Regardless of how the subsidy is funded, however, someone in the domestic economy must ultimately pay for it.

Export subsidy effects on the exporting country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers and producers. The net effect consists of three components: a negative terms of trade effect ($f + g + h$), a negative consumption distortion (b), and a negative production distortion (d).

Since all three components are negative, the export subsidy must result in a reduction in national welfare for the exporting country. However, it is important to note that a redistribution of income occurs—that is, some groups gain while others lose. The likely reason governments implement export subsidies is because they will benefit domestic exporting firms. The concerns of consumers must be weighed less heavily in their calculation since the sum of their losses exceeds the sum of the producers' gains.

Export subsidy effects on the importing country's consumers. Consumers of the product in the importing country experience an increase in well-being as a result of the export subsidy. The decrease in the price of both imported goods and the domestic substitutes increases the amount of consumer surplus in the market.

Export subsidy effects on the importing country's producers. Producers in the importing country suffer a decrease in well-being as a result of the export subsidy. The decrease in the price of their product on the domestic market reduces producer surplus in the industry. The price decrease also induces a decrease in the output of existing firms, a decrease in employment, and a decrease in profit, payments, or both to fixed costs.

Export subsidy effects on the importing country's government. There is no effect on the importing country's government revenue as a result of the exporter's subsidy.

Export subsidy effects on the importing country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the government. The net effect consists of three components: a positive terms of trade effect ($F + G + H$), a negative production distortion (F), and a negative consumption distortion (H).

Although there are both positive and negative elements, the net national welfare effect reduces to area G , which is positive. This means that an export subsidy

implemented by a large exporting country in a perfectly competitive market *will raise national welfare in the importing country.*

This result has inspired some economists to argue that the proper response for an importing country when its trading partner implements an export subsidy is simply to send along a thank you note.

It is worth noting here that the World Trade Organization (WTO) allows countries to impose countervailing duties to retaliate against its trading partners when it can be shown that an exporting country's government has used export subsidies.

It is also important to note that not everyone's welfare rises when there is an increase in national welfare. Instead, there is a redistribution of income. Consumers of the product will benefit, but producers and payers of government taxes will lose. A national welfare increase, then, means that the sum of the gains exceeds the sum of the losses across all individuals in the economy. Economists generally argue that, in this case, compensation from winners to losers can potentially alleviate the redistribution problem.

Export subsidy effects on world welfare. The effect on world welfare is found by summing the national welfare effects on the importing and exporting countries. By noting that the terms of trade gain to the exporter is equal to the terms of trade loss to the importer, the world welfare effect reduces to four components: the importer's negative production distortion (B), the importer's negative consumption distortion (D), the exporter's negative consumption distortion (f), and the exporter's negative production distortion (h). Since each of these is negative, the world welfare effect of the export subsidy is negative. The sum of the losses in the world exceeds the sum of the gains. In other words, we can say that an export subsidy results in a reduction in world production and consumption efficiency.

KEY TAKEAWAYS

- An export subsidy lowers consumer surplus and raises producer surplus in the exporter market.
- An export subsidy raises producer surplus in the export market and lowers it in the import country market.
- National welfare falls when a large country implements an export subsidy.
- National welfare in the importing country rises when a large exporting country implements an export subsidy.
- An export subsidy of any size will reduce world production and consumption efficiency and thus cause world welfare to fall.

EXERCISE

1. Consider the following trade policy action (applied by the domestic country) listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of the policy on the variables listed in the first column:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

Use a partial equilibrium model to determine the answers and assume that the shapes of the supply and demand curves are “normal.” Assume that the policy does not begin with, or result in, prohibitive trade policies. Also assume that the policy does not correct for market imperfections or distortions.

For example, an export subsidy applied by a large country will cause an increase in the domestic price of the export good; therefore a + is placed in the first box of the table.

TABLE 7.11 EXPORT SUBSIDY EFFECTS

	Export Subsidy by a Large Country
Domestic Market Price	+
Domestic Industry Employment	
Domestic Consumer Welfare	
Domestic Producer Welfare	

	Export Subsidy by a Large Country
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	
Foreign Consumer Welfare	
Foreign Producer Welfare	
Foreign National Welfare	

7.18 Countervailing Duties

LEARNING OBJECTIVES

1. Understand the basic mechanics of an antisubsidy law allowable to members of the World Trade Organization (WTO).
2. Identify the effects of a countervailing duty by an import country in response to a foreign government export subsidy.

The World Trade Organization (WTO) allows countries to implement antisubsidy legislation. The law allows a country to place a **countervailing duty (CVD)**¹¹ on imports when a foreign government subsidizes exports of the product, which in turn causes injury to the import-competing firms. The countervailing duty is a tariff designed to “counter” the effects of the foreign export subsidy. The purpose of this section is to explain the effects of a countervailing duty in a perfectly competitive market setting. See [Chapter 1 "Introductory Trade Issues: History, Institutions, and Legal Framework"](#), [Section 1.5 "The General Agreement on Tariffs and Trade \(GATT\)"](#) for a more complete description of the antisubsidy law.

We will assume that there are two large countries trading a particular product in a partial equilibrium model. The exporting country initially sets a specific export subsidy. That action is countered with a CVD implemented by the importing country. We will first describe the effects of the export subsidy (which will closely mimic the analysis in [Chapter 7 "Trade Policy Effects with Perfectly Competitive Markets"](#), [Section 7.16 "Export Subsidies: Large Country Price Effects"](#) and [Chapter 7 "Trade Policy Effects with Perfectly Competitive Markets"](#), [Section 7.17 "Export Subsidies: Large Country Welfare Effects"](#)), after which we will consider the effects of the CVD action in response.

The Initial Export Subsidy

An export subsidy will reduce the price of the good in the import market and raise the price of the good in the export market relative to the free trade price. After the subsidy is imposed, the following two conditions will describe the new equilibrium:

11. A tariff on a product levied against a country that subsidizes the export of that product and that causes injury to the import-competing industry.

$$P_S^{EX} = P_S^{IM} + S$$

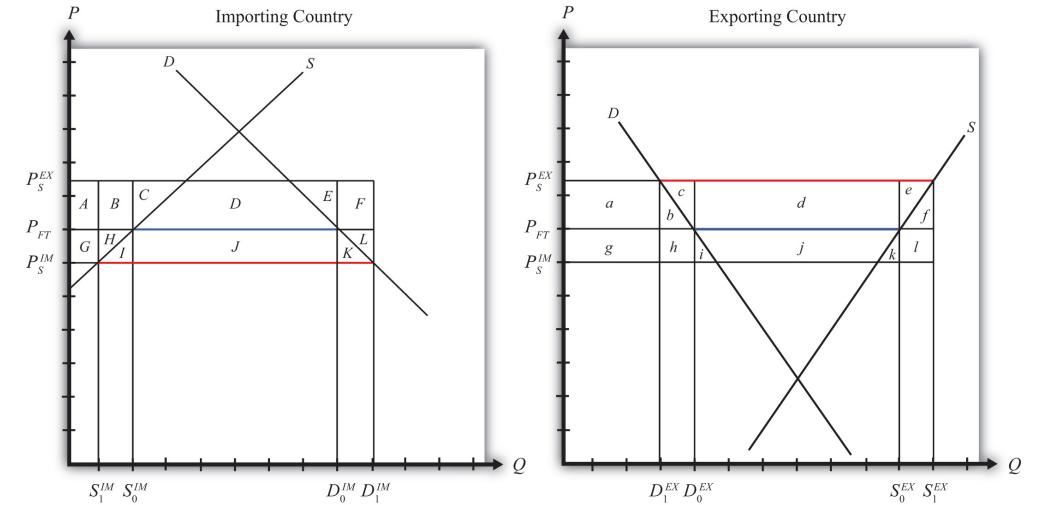
and

$$XS(P_S^{EX}) = MD(P_S^{IM}),$$

where S is the specific export subsidy, P_S^{IM} is the price that prevails in the import market after the subsidy, and P_S^{EX} is the price that prevails in the export market after the subsidy. The first condition means that prices in the two countries must differ by the amount of the subsidy. The second condition means that export supply at the price that now prevails in the export market must equal import demand at the price that prevails in the import market.

The effects of the subsidy are depicted in [Figure 7.33 "Depicting an Export Subsidy and a CVD"](#). The initial free trade price is labeled P_{FT} . In free trade, the exporting country exports ($S_0^{EX} - D_0^{EX}$) and the importing country imports ($D_0^{IM} - S_0^{IM}$). Since there are only two countries in the model, free trade exports are equal to imports and are shown as the blue line segments in the diagram. When the subsidy is imposed, the price in the export market rises to P_S^{EX} , while the price in the import market falls to P_S^{IM} . The higher level of exports with the subsidy, given by ($S_1^{EX} - D_1^{EX}$), is equal to imports, given by ($D_1^{IM} - S_1^{IM}$), and is depicted by the red line segments in [Figure 7.30 "Effects of a World Price Decrease"](#).

Figure 7.33 Depicting an Export Subsidy and a CVD



[Table 7.12 "Welfare Effects of the Initial Export Subsidy"](#) provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the governments in the importing and exporting countries as a result of the subsidy. The aggregate national welfare effects and the world welfare effects are also shown.

Table 7.12 Welfare Effects of the Initial Export Subsidy

	Importing Country	Exporting Country
Consumer Surplus	+ (G + H + I + J + K)	- (a + b)
Producer Surplus	- (G + H)	+ (a + b + c + d + e)
Govt. Revenue	0	- (b + c + d + e + f + h + i + j + k + l)
National Welfare	+ I + J + K	- (b + f + h + i + j + k + l)
World Welfare	- (I + K) - (b + f)	

Table 7.12 "Welfare Effects of the Initial Export Subsidy" shows that in the case of a large exporting country, the export producers benefit from the subsidy, while the consumers of the product in the exporting country lose. Because of the cost of the subsidy to the exporting country government, which must ultimately be paid by the taxpayers, the net national welfare effect for the exporting country is negative.

The importing country also experiences an income redistribution. The consumers in the importing country benefit from the foreign subsidy, while import-competing producers suffer losses. The net effect for the importing country is positive since the gains to consumers outweigh the losses to producers.

The world welfare effects of the export subsidy are also negative.

The Countervailing Duty

Despite the fact that the export subsidy generates net benefits for the importing country, the importing country is allowed under WTO rules to protect itself from these benefits. A CVD may be placed if it can be shown that a subsidy is indeed in place and if the subsidy causes injury to the import-competing firms.

It is worth emphasizing that the antisubsidy law, in this case, does not protect the “country,” nor does it protect consumers. The law is designed to aid import firms exclusively. No evaluation of the effects on consumers and no evaluation of the national welfare effects are required by the law. The only requirement is that injury be caused to the import-competing firms.

In this simple example of a large country implementing an export subsidy, injury would indeed be apparent. The export subsidy lowers the price of the good in the import market in this model and causes an increase in imports from abroad. Supply

by the import-competing firms would fall (from S_0^{IM} to S_1^{IM} in [Figure 7.33 "Depicting an Export Subsidy and a CVD"](#)). Producer surplus, indicating a reduction in industry profits, would also fall. Since less output would be produced by the import-competing industry, the industry would need fewer factors of production. This would likely mean a reduction in the number of workers employed in the industry. In the adjustment process, firms in the industry may lay off workers and close factories. All these effects are valid criteria used to judge injury in CVD cases.

So let's consider the effects of a countervailing duty in response to the export subsidy described above. A CVD is simply a tariff set on imports to counter the effects of the foreign export subsidy. CVD laws require that the size of the CVD be just enough to offset the effects of the export subsidy. In the United States, the U.S. International Trade Administration determines the size of the foreign subsidy. If a CVD action is taken, the CVD is set equal to the foreign subsidy.

So imagine that the importing country now sets a specific CVD (t) equal to the original export subsidy (S). As with any tariff set by a large importing country, the tariff will cause the price in the importing country to rise and the price in the exporting country to fall. What's different from the standard tariff analysis is that the prices in this case are not equal to each other. Instead, the price in the import market begins lower—by the amount of the export subsidy, S —than the price in the export market. The CVD, then, will drive the prices in the two markets back together.

The final equilibrium must satisfy the following two conditions:

$$P_{S+t}^{EX} + t = P_{S+t}^{IM} + S$$

and

$$XS(P_{S+t}^{EX}) = MD(P_{S+t}^{IM}).$$

However, since $t = S$, the first condition reduces to $P_{S+t}^{EX} = P_{S+t}^{IM}$. This means that in the final equilibrium, the prices must be equal in both countries and export supply must be equal to import demand. These conditions are satisfied only at the free trade price.

Thus the effect of the CVD is to force the prices in the two markets back to the free trade prices.

As a result, imports will fall in the importing country (back to $D_0^{IM} - S_0^{IM}$ in [Figure 7.33 "Depicting an Export Subsidy and a CVD"](#)), domestic supply will rise (from S_1^{IM} to S_0^{IM}), employment in the import-competing industry will rise, and producer surplus in the industry will also rise. Thus the CVD will be effective in eliminating the injury caused to import-competing firms.

Welfare Effects of the CVD

But let's also take a look at the overall welfare effects of the CVD, assuming, as is often the case, that the CVD and the export subsidy remain in place. There are two ways to consider the effects of the CVD. We can look at the effects relative to when just the export subsidy was in place. Or we can look at the effects relative to when there was no export subsidy and no CVD. We'll do it both ways.

First, let's consider the welfare effects of the CVD relative to when the export subsidy alone was in place. These effects are summarized in [Table 7.13 "Welfare Effects of a CVD"](#).

Table 7.13 Welfare Effects of a CVD

	Importing Country	Exporting Country
Consumer Surplus	$-(G + H + I + J + K)$	$+(a + b)$
Producer Surplus	$+(G + H)$	$-(a + b + c + d + e)$
Govt. Revenue	$+(C + D + E + J)$	$+(b + c + e + f + h + l)$
National Welfare	$+(C + D + E) - (I + K)$	$+(b + f + h + l) - (d)$
World Welfare	$+(b + f + h + l) - (I + K) =$	$b + f + I + K$

Note that the effects on consumers and producers in both countries are equal and opposite to the effects of the export subsidy. Thus producers in the import-competing industry gain in surplus from the CVD exactly what they had lost as a result of the foreign export subsidy. Consumers in the import industry lose from the CVD, producers in the exporting country lose, and consumers in the exporting country gain.

The importing government now collects tariff revenue from the CVD, which benefits someone in the importing country. The exporting government, however, experiences a reduction in its subsidy expenditures. This occurs because the CVD

reduces trade and thus reduces the number of units exported. As a result, the government (i.e., the taxpayers) in the exporting country benefits from the CVD.

The national welfare effects in both countries are ambiguous in general. In the importing country, a terms of trade gain may outweigh two deadweight losses and cause national welfare to rise even further. Interestingly, the export subsidy and the CVD may each raise welfare for the importing country. In the export country, the net national welfare effect may be positive or negative.

The world welfare effects are found by summing the national welfare effects on both countries. The expression is simplified first by noting that area $(C + D + E)$ = area (d) and second by noting that area (h) = twice area I , or $(2I)$, and area (l) = area $(2K)$. The final expression shows that world welfare will rise as a result of the CVD.

Welfare Effects of the Combined Policies (Export Subsidy plus CVD)

Next, let's consider the welfare effects of the export subsidy and the CVD combined. In this case, we compare the welfare status of each country after both policies are in place relative to when neither policy is imposed. The effects can be calculated either by summing the individual welfare effects of each of the two stages depicted above or by noting that prices have not changed from the initial presubsidy state to the final post-CVD state but that the governments do have expenditures and receipts, respectively.

The welfare effects are summarized in [Table 7.14 "Welfare Effects of an Export Subsidy plus a CVD"](#).

Table 7.14 Welfare Effects of an Export Subsidy plus a CVD

	Importing Country	Exporting Country
Consumer Surplus	0	0
Producer Surplus	0	0
Govt. Revenue	$+ (C + D + E + J)$	$- (d + i + j + k)$
National Welfare	$+ (C + D + E + J)$	$- (d + i + j + k)$
World Welfare	0	

Since the prices in each country after the CVD are the same as prices before the export subsidy, there is ultimately no change in producer or consumer surplus in either country. Everyone participating in the market is left as well off as they were at the start.

However, since the exporting country maintains the export subsidy and the import country maintains the CVD, there are government revenue effects. In the exporting country, the government continues to make expenditures for the export subsidy. This represents a cost to the country's taxpayers that does not even generate the intended benefit for the export industry. In the importing country, the government collects tariff revenue as a result of the CVD. This generates benefits to the recipients of the resulting additional government spending.

The net national welfare effect in each country is the same as the government effects. This means that the importing country benefits from the export subsidy plus CVD, while the exporting country loses from the combined policies.

The world welfare effect of the combined policies is neutral. This means that the exporting country loses exactly the same amount as the importing country gains. The ultimate effect of the export subsidy plus the CVD is that the exporting country's government transfers money to the importing country's government with consumers and producers left unaffected. In practice, exporting country producers receive an export subsidy payment from their government when their product leaves the port bound for the importing country. When the product arrives, the importing country's government collects a tariff (or a CVD) exactly equal to the subsidy payment. Thus the export firms turn over the extra monies they had just received from their own government to the government of the importing country.

These effects described here hold only for markets that are perfectly competitive. If the markets are oligopolistic, or contain market imperfections or other distortions, then the effects of the export subsidy and CVD may differ.

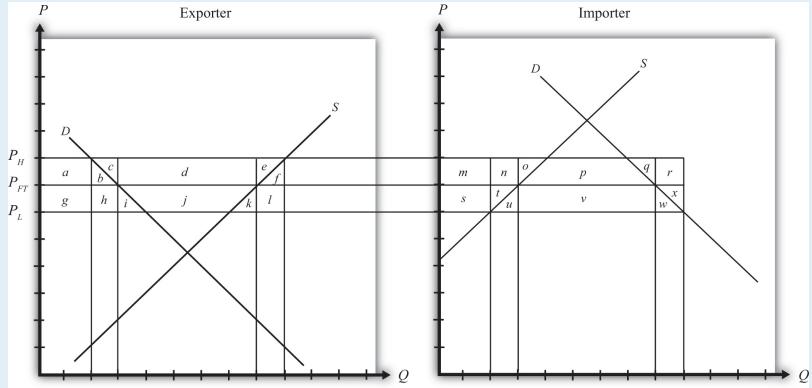
KEY TAKEAWAYS

- An antisubsidy law, allowable under the WTO agreement, enables countries to apply a countervailing duty (CVD)—that is, an import tariff—equal in value to the export subsidy that is shown to be in place by the exporting country in a particular product market.
- A CVD will cause the price of the product in both countries to revert to the free trade price.
- A CVD will raise producer surplus and lower consumer surplus in the import country relative to the equilibrium with just the export subsidy in place.
- The net effect of a CVD and the foreign export subsidy together is a transfer of income from the export country’s government to the import country’s government.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe the tariff used to offset the injurious effects of a foreign government export subsidy.
 - b. Of *increase, decrease, or stay the same*, the effect on the domestic price of a product, relative to free trade, if a domestic exporting country has an export subsidy of S and the foreign importer has a $CVD = C$ in place such that $C = S$.
 - c. Of *increase, decrease, or stay the same*, the effect on domestic production of a product, relative to free trade, if a domestic exporting country has an export subsidy of S and the foreign importer has a $CVD = C$ in place such that $C = S$.
 - d. Of *increase, decrease, or stay the same*, the effect on domestic consumption of a product, relative to free trade, if a domestic exporting country has an export subsidy of S and the foreign importer has a $CVD = C$ in place such that $C = S$.
 - e. Of *increase, decrease, or stay the same*, the effect on the domestic price of a product, relative to free trade, if a domestic exporting country has an export subsidy of S and the foreign importer has a $CVD = C$ in place such that $C > S$.
2. Consider a market for computers in two large countries. Suppose the exporting country imposes a specific export subsidy equal to $P_H - P_L$. Afterward, the importing country retaliates with a countervailing duty also set equal to $P_H - P_L$. Use the diagram below to answer the following questions.

Figure 7.34
Two Large Trading Countries



- What is the change in consumer surplus in the exporting country when the export subsidy is imposed?
- What is the change in producer surplus in the exporting country when the export subsidy is imposed?
- What are government subsidy payments in the exporting country when the export subsidy is imposed?
- What is the net national welfare effect in the exporting country when the export subsidy is imposed?
- What is the net national welfare effect in the importing country when the subsidy is imposed?
- What is the change in consumer surplus in the importing country (relative to the subsidy in place) with the CVD?
- What is the change in producer surplus in the importing country (relative to the subsidy in place) with the CVD?
- What is the change in government revenue in the importing country (relative to the subsidy in place) with the CVD?
- What is the change in government revenue in the exporting country (relative to the subsidy in place) with the CVD?
- What condition must hold for the CVD to improve welfare in the importing country (relative to the subsidy)?

7.19 Voluntary Export Restraints (VERs): Large Country Price Effects

LEARNING OBJECTIVES

1. Identify the effects of a voluntary export restraint, or export quota, on prices in both countries and the quantity traded.
2. Know the equilibrium conditions that must prevail in a voluntary export restraint (VER) equilibrium.

Suppose the United States, an exporting country in free trade, imposes a binding export quota, often called a voluntary export restraint (VER) when implemented bilaterally, on wheat exports to Mexico. The VER will restrict the flow of wheat across the border. Since the United States is a large exporter, the supply of wheat to the Mexican market will fall, and if the price remained the same it would cause excess demand for wheat in the market. The excess demand will induce an increase in the price of wheat. Since wheat is homogeneous and the market is perfectly competitive, the price of all wheat sold in Mexico, both Mexican wheat and U.S. imports, will rise in price. The higher price will, in turn, reduce demand and increase domestic supply, causing a reduction in Mexico's import demand.

The restricted wheat supply to Mexico will shift supply back to the U.S. market, causing excess supply in the U.S. market at the original price and a reduction in the U.S. price. The lower price will, in turn, reduce U.S. supply, raise U.S. demand, and cause a reduction in U.S. export supply.

These price effects are identical in direction to the price effects of an import tax and an import quota by the importer country, and an export tax by the exporting country.

A new VER equilibrium will be reached when the following two conditions are satisfied:

$$MD^{Mex}(P_V^{Mex}) = \bar{Q}$$

and

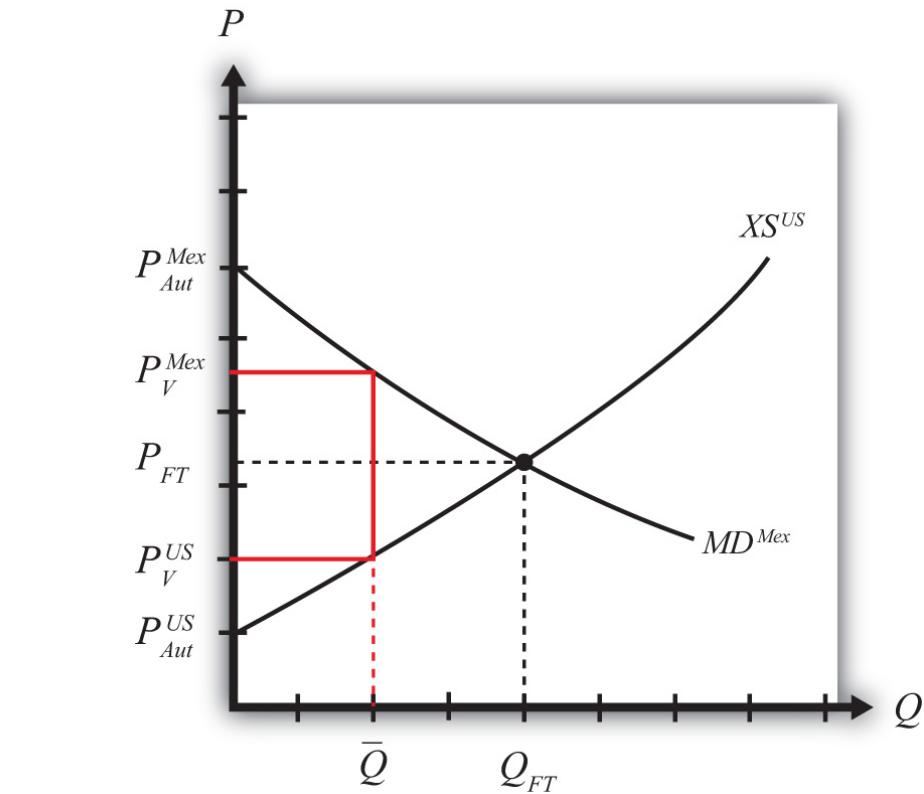
$$XS^{US}(P_V^{US}) = \bar{Q},$$

where \bar{Q} is the quantity at which the VER is set, P_V^{Mex} is the price in Mexico after the VER, and P_V^{US} is the price in the United States after the VER.

The first condition says that the price must change in Mexico such that import demand falls to the VER level \bar{Q} . In order for this to occur, the price in Mexico rises. The second condition says that the price must change in the United States such that export supply falls to the VER level \bar{Q} . In order for this to occur, the price in the United States falls.

The VER equilibrium is depicted graphically in [Figure 7.35 "Depicting a VER Equilibrium: Large Country Case"](#). The Mexican price of wheat rises from P_{FT} to P_V^{Mex} , which is sufficient to reduce its import demand from Q_{FT} to \bar{Q} . The U.S. price of wheat falls from P_{FT} to P_V^{US} , which is sufficient to reduce its export supply also from Q_{FT} to \bar{Q} .

Figure 7.35 Depicting a VER Equilibrium: Large Country Case



Notice that a unique set of prices satisfies the equilibrium conditions for every potential VER that is set. If the VER were set lower than \bar{Q} , the price wedge would

rise, causing a further increase in the Mexican price and a further decrease in the U.S. price.

At the extreme, if the VER were set equal to zero, then the prices in each country would revert to their autarky levels. In this case, the VER would *prohibit* trade. This situation is similar to an export embargo.

KEY TAKEAWAYS

- A VER implemented by an exporting country will reduce the domestic price and, in the case of a large country, raise the foreign price.
- The difference between the domestic and foreign price after the VER represents a quota rent.
- A VER will reduce the quantity of exports to the quota amount.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The direction of change of the domestic price after a binding VER is implemented by an exporting country.
 - b. The direction of change of the foreign importer price after a binding VER is implemented by a large exporting country.
 - c. Of *increase, decrease, or stay the same*, this is the effect on the domestic price after a nonbinding export quota is implemented by an exporting country.
 - d. Of *increase, decrease, or stay the same*, this is the effect on the quantity of wheat exports if a binding VER is implemented.
 - e. Of *increase, decrease, or stay the same*, this is the effect on foreign imports of shoes if a binding VER is implemented by an exporting country.

7.20 Administration of a Voluntary Export Restraint

LEARNING OBJECTIVE

1. Learn the ways in which a voluntary export restraint (VER) can be implemented to monitor and assure that only the specified amount is exported to the targeted country.

When a government sets a quantity restriction, the government must implement procedures to prevent exports beyond the restricted level. A binding voluntary export restraint (VER) will result in a higher price in the import country and in the case of a large country, a reduction in the price in the exporter's market. The price wedge would generate profit opportunities for anyone who could purchase (or produce) the product at the lower price (or cost) in the export market and resell it at the higher price in the import market.

Three basic methods are used to administer VERs.

1. *Offer export rights on a first-come, first-served basis.* The government could allow exports to exit freely from the start of the year until the VER limit is reached. Once filled, customs officials would prohibit export of the product for the remainder of the year. If administered in this way, the VER may result in a fluctuating price for the product over the year. During the open period, a sufficient amount of imports may flow in to achieve free trade prices. Once the window is closed, prices would revert to the autarky prices.
2. *Auction export rights.* Essentially the government could sell quota tickets where each ticket presented to a customs official would allow the exit of one unit of the good. If the tickets are auctioned, or if the price is determined competitively, the price at which each ticket would be sold is the difference in prices that exist between the export and import market. The holder of a quota ticket can buy the product at the low price in the exporter's market and resell it at the higher price in the importer's market. If there are no transportation costs, a quota holder can make a pure profit, called a quota rent, equal to the difference in prices. If the government sells the quota tickets at the maximum attainable price, then the government would receive all the quota rents.

3. *Give away export rights.* The government could give away the export rights by allocating quota tickets to appropriate individuals. The recipient of a quota ticket essentially receives a windfall profit since, in the absence of transportation costs, they can claim the entire quota rent at no cost to themselves. Many times governments allocate the quota tickets to domestic exporting companies based on past market shares. Thus, if an exporter had exported 40 percent of all exports before the VER, then it would be given 40 percent of the quota tickets. It is worth noting that because quota rents are so valuable, a government can use them to direct rents toward its political supporters.

KEY TAKEAWAYS

- To administer a VER, countries generally assign export rights, or licenses, with the allowable import quantity limited in total to quota level.
- The government earns revenue from the quota rents if it allocates the export licenses via auction or sale.
- If the government gives the export rights away, as it typically does in these cases, the recipients of the rights, typically the export firms themselves, earn the quota rents.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *domestic* or *foreign* residents, this group receives quota rents when the government sells the right to export.
 - b. The term for the quota allocation method in which exports are allowed until the quota limit is reached.
 - c. The term used to describe the sale of quota rights to the highest bidder.
 - d. The likely recipients if new quota rights are given away by the government.
 - e. The term used to describe the profit made by a quota rights holder who can purchase the product cheaper in the export market and sell it for more in the import market.

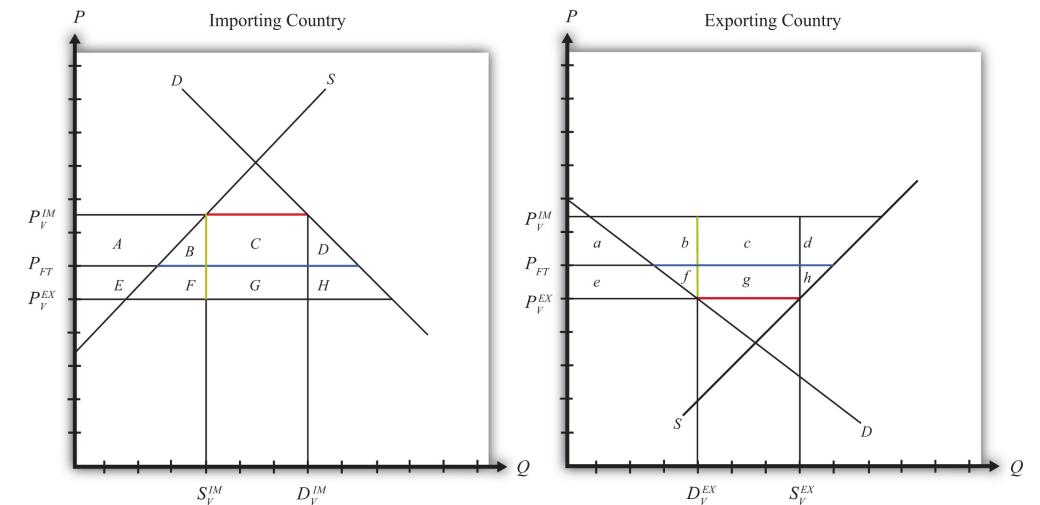
7.21 Voluntary Export Restraints: Large Country Welfare Effects

LEARNING OBJECTIVES

1. Use a partial equilibrium diagram to identify the welfare effects of a voluntary export restraint (VER) on producer and consumer groups and the government in the exporting and importing countries.
2. Calculate the national and world welfare effects of a VER in the case of a large country.

Suppose for simplicity that there are only two trading countries: one importing country and one exporting country. The supply and demand curves for the two countries are shown in [Figure 7.36 "Welfare Effects of a VER: Large Country Case"](#). P_{FT} is the free trade equilibrium price. At that price, the excess demand by the importing country equals excess supply by the exporter.

Figure 7.36 Welfare Effects of a VER: Large Country Case



The quantity of imports and exports is shown as the blue line segment on each country's graph (the horizontal distance between the supply and demand curves at the free trade price). Suppose the large exporting country implements a binding voluntary export restraint set equal to the length of the red line segment. When a new equilibrium is reached, the price in the importing country will rise to the level at which import demand is equal to the quota level. The price in the exporting country will fall until export supply is equal to the quota level.

Table 7.15 "Welfare Effects of a Voluntary Export Restraint" provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the governments in the importing and exporting countries. The aggregate national welfare effects and the world welfare effects are also shown.

Table 7.15 Welfare Effects of a Voluntary Export Restraint

	Importing Country	Exporting Country
Consumer Surplus	$-(A + B + C + D)$	$+e$
Producer Surplus	$+A$	$-(e + f + g + h)$
Quota Rents	0	$+(c + g)$
National Welfare	$-(B + C + D)$	$c - (f + h)$
World Welfare	$-(B + D) - (f + h)$	

Refer to [Table 7.15 "Welfare Effects of a Voluntary Export Restraint"](#) and [Figure 7.36 "Welfare Effects of a VER: Large Country Case"](#) to see how the magnitudes of the changes are represented.

VER effects on the exporting country's consumers. Consumers of the product in the exporting country experience an increase in well-being as a result of the VER. The decrease in their domestic price raises the amount of consumer surplus in the market.

VER effects on the exporting country's producers. Producers in the exporting country experience a decrease in well-being as a result of the quota. The decrease in the price of their product in their own market decreases producer surplus in the industry. The price decline also induces a decrease in output, a decrease in employment, and a decrease in profit, payments, or both to fixed costs.

VER effects on the quota rents. Who receives the quota rents depends on how the government administers the quota.

1. If the government auctions the quota rights for their full price, then the government receives the quota rents. In this case, the quota is equivalent to a specific export tax set equal to the difference in prices ($T = P_V^M - P_V^{EX}$), shown as the length of the green line segment in [Figure 7.36 "Welfare Effects of a VER: Large Country Case"](#).

2. If the government gives away the quota rights, then the quota rents accrue to whoever receives these rights. Typically, they would be given to the exporting producers, which would serve to offset the producer surplus losses. It is conceivable that the quota rents may exceed the surplus loss so that the export industry is better off with the VER than without. Regardless, the benefits would remain in the domestic economy.

VER effects on the exporting country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the recipients of the quota rents. The net effect consists of three components: a positive terms of trade effect (c), a negative production distortion (h), and a negative consumption distortion (f).

Because there are both positive and negative elements, the net national welfare effect can be either positive or negative. The interesting result, however, is that it can be *positive*. This means that a VER implemented by a large exporting country *may* raise national welfare.

Generally speaking, the following are true:

1. Whenever a large country implements a small restriction on exports, it will raise national welfare.
2. If the VER is too restrictive, national welfare will fall.
3. There will be a positive quota level that will maximize national welfare.

However, it is also important to note that not everyone's welfare rises when there is an increase in national welfare. Instead, there is a redistribution of income. Consumers of the product and recipients of the quota rents will benefit, but producers may lose. A national welfare increase, then, means that the sum of the gains exceeds the sum of the losses across all individuals in the economy. Economists generally argue that, in this case, compensation from winners to losers can potentially alleviate the redistribution problem.

VER effects on the importing country's consumers. Consumers of the product in the importing country suffer a reduction in well-being as a result of the VER. The increase in the domestic price of both imported goods and the domestic substitutes reduces the amount of consumer surplus in the market.

VER effects on the importing country's producers. Producers in the importing country experience an increase in well-being as a result of the VER. The increase in the price

of their product increases producer surplus in the industry. The price increases also induce an increase in the output of existing firms (and perhaps the addition of new firms), an increase in employment, and an increase in profit, payments, or both to fixed costs.

VER effects on the importing country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers and producers. The net effect consists of three components: a negative terms of trade effect (C), a negative consumption distortion (D), and a negative production distortion (B).

Since all three components are negative, the VER must result in a reduction in national welfare for the importing country. However, it is important to note that a redistribution of income occurs—that is, some groups gain while others lose. This is especially important because VERs are often suggested by the importing country. This occurs because the importing country’s government is pressured by the import-competing producers to provide protection in the form of an import tariff or quota. Government reluctance to use these policies often leads the importer to negotiate VERs with the exporting country. Although the importing country’s national welfare is reduced, the import-competing producers gain nonetheless.

VER effects on world welfare. The effect on world welfare is found by summing the national welfare effects on the importing and exporting countries. By noting that the terms of trade gain to the importer is equal to the terms of trade loss to the exporter, the world welfare effect reduces to four components: the importer’s negative production distortion (B), the importer’s negative consumption distortion (D), the exporter’s negative consumption distortion (f), and the exporter’s negative production distortion (h). Since each of these is negative, the world welfare effect of the VER is negative. The sum of the losses in the world exceeds the sum of the gains. In other words, we can say that a VER results in a reduction in world production and consumption efficiency.

KEY TAKEAWAYS

- A VER raises consumer surplus in the export market and lowers it in the import country market.
- A VER lowers producer surplus in the export market and raises it in the import country market.
- National welfare may rise or fall when a large exporting country implements a VER.
- National welfare in the importing country rises when a large exporting country implements a VER.
- A VER of any size will reduce world production and consumption efficiency and thus cause world welfare to fall.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The direction of change of domestic producer surplus when a binding VER is implemented by an exporting country.
 - b. The direction of change of foreign producer surplus when a binding VER is implemented by an exporting country.
 - c. The direction of change of domestic consumer surplus when a binding VER is implemented by an exporting country.
 - d. The direction of change of foreign consumer surplus when a binding VER is implemented by an exporting country.
2. Consider the following trade policy action listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of the policy on the variables listed in the first column:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

Use a partial equilibrium model to determine the answers, and assume that the shapes of the supply and demand curves are “normal.” Assume that the policy does not begin with, or result in, prohibitive trade policies. Also assume that the policy does not correct for market imperfections or distortions.

TABLE 7.16 EFFECTS OF A VER ELIMINATION

	Elimination of a Binding VER by a Large Exporting Country
Domestic Market Price	
Domestic Industry Employment	
Domestic Consumer Welfare	
Domestic Producer Welfare	
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	
Foreign Consumer Welfare	
Foreign Producer Welfare	
Foreign National Welfare	

7.22 Export Taxes: Large Country Price Effects

LEARNING OBJECTIVES

1. Identify the effects of an export tax on prices in both countries and the quantity traded in the case of a large country.
2. Know the equilibrium conditions that must prevail in an export tax equilibrium.

Suppose the United States, the exporting country in free trade, imposes a specific export tax on exports of wheat. A tax on exports will reduce the flow of wheat across the border. It will now cost more to move the product from the United States into Mexico.

As a result, the supply of wheat to the Mexican market will fall, inducing an increase in the price of wheat. Since the United States is assumed to be a large country, the price of all wheat sold in Mexico, both Mexican wheat and U.S. imports, will rise in price. The higher price will reduce Mexico's import demand.

The reduced wheat supply to Mexico will shift supply back to the U.S. market and induce a reduction in the U.S. price. The lower price will reduce U.S. export supply.

These price effects are identical in direction to the price effects of a tariff, an import quota, and a voluntary export restraint.

A new tax-ridden equilibrium will be reached when the following two conditions are satisfied:

$$P_T^{Mex} = P_T^{US} + T$$

and

$$XS^{US}(P_T^{US}) = MD^{Mex}(P_T^{Mex}),$$

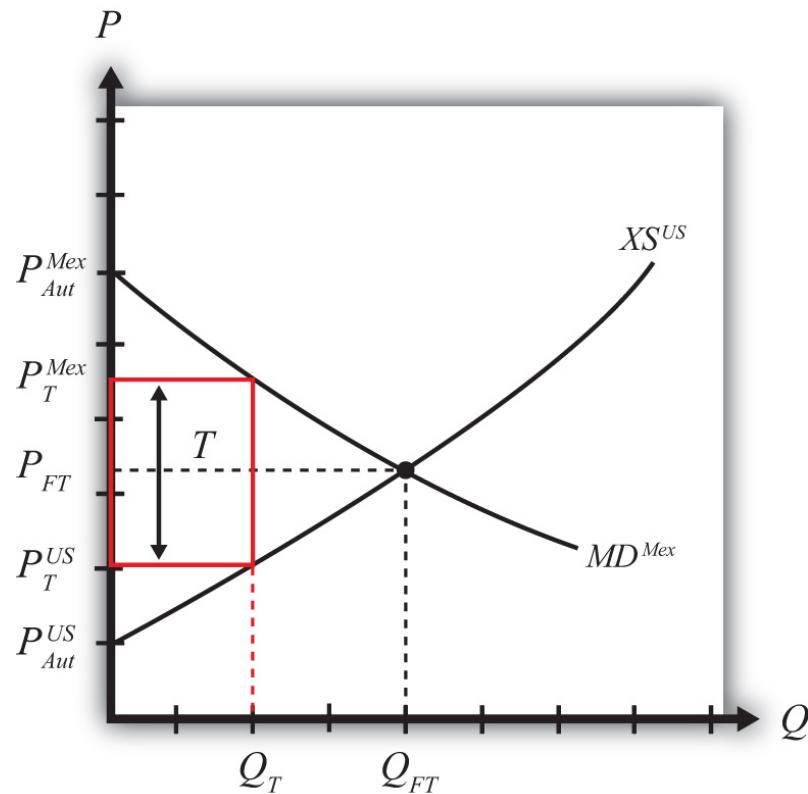
where T is the export tax, P_T^{Mex} is the price in Mexico after the tax, and P_T^{US} is the price in the United States after the tax.

The first condition represents a price wedge between the final U.S. price and the Mexican price equal to the amount of the export tax. The prices must differ from the tax because U.S. suppliers of wheat must receive the same price for their product regardless of whether the product is sold in the United States or Mexico, and all wheat sold in Mexico must be sold at the same price. Since a tax is collected at the border, the only way for these price equalities within countries to arise is if the price differs across countries by the amount of the tax.

The second condition states that the amount the United States wants to export at its new lower price must be equal to the amount Mexico wants to import at its new higher price. This condition guarantees that world supply of wheat equals world demand for wheat.

The export tax equilibrium is depicted graphically in [Figure 7.37 "Depicting an Export Tax Equilibrium: Large Country Case"](#). The Mexican price of wheat rises from P_{FT} to P_T^{Mex} , which reduces its import demand from Q_{FT} to Q_T . The U.S. price of wheat falls from P_{FT} to P_T^{US} , which reduces its export supply also from Q_{FT} to Q_T . The difference in the prices between the two markets is equal to the export tax rate T .

Figure 7.37 Depicting an Export Tax Equilibrium: Large Country Case



Notice that there is a unique set of prices that satisfies the equilibrium conditions for every potential export tax that is set. If the tax was set higher than T , the price wedge would rise, causing a further increase in the Mexican price, a further decrease in the U.S. price, and a further reduction in the quantity traded.

KEY TAKEAWAYS

- An export tax will lower the domestic price and, in the case of a large country, raise the foreign price.
- An export tax will decrease the quantity of exports.
- The export tax will drive a price wedge, equal to the tax rate, between the domestic price and the foreign price of the product.
- With the export tax in place in a two-country model, export supply at the lower domestic price will equal import demand at the higher foreign price.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The kind of power a country is said to have when its exports make up a significant share of the world market.
 - b. The direction of change of the domestic price after an export tax is implemented by a domestic country.
 - c. The direction of change of the foreign price after an export tax is implemented by a large domestic country.
 - d. The price of tea in the exporting country if the exporter sets an export tax of \$0.75 per pound and if the importer country price is \$4.75 inclusive of the tax.
 - e. Of *increase, decrease, or stay the same*, this is the effect on exports of wheat if an export tax on wheat is implemented.
 - f. Of *increase, decrease, or stay the same*, this is the effect on foreign imports of wheat if an export tax on wheat is implemented by an exporting country.

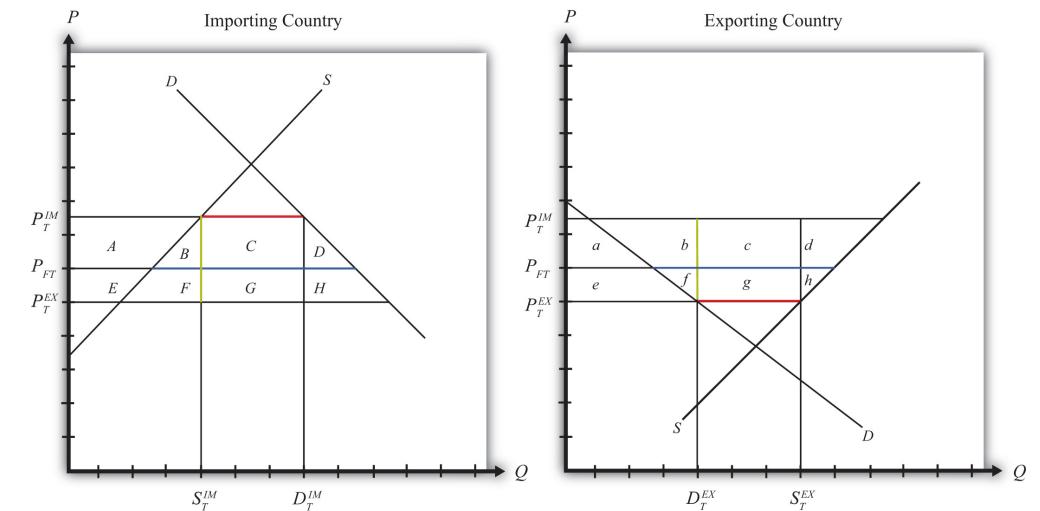
7.23 Export Taxes: Large Country Welfare Effects

LEARNING OBJECTIVES

1. Use a partial equilibrium diagram to identify the welfare effects of an export tax on producer and consumer groups and the government in the exporting and importing countries.
2. Calculate the national and world welfare effects of an export tax.

Suppose that there are only two trading countries: one importing country and one exporting country. The supply and demand curves for the two countries are shown in [Figure 7.38 "Welfare Effects of an Export Tax: Large Country Case"](#). P_{FT} is the free trade equilibrium price. At that price, the excess demand by the importing country equals excess supply by the exporter.

Figure 7.38 Welfare Effects of an Export Tax: Large Country Case



The quantity of imports and exports is shown as the blue line segment on each country's graph (the horizontal distance between the supply and demand curves at the free trade price). When a large exporting country implements an export tax, it will cause a decrease in the price of the good on the domestic market and an increase in the price in the rest of the world (RoW). Suppose after the tax, the price in the importing country rises to P_T^{IM} and the price in the exporting country falls to P_T^{EX} . If the tax is a specific tax, then the tax rate would be $T = P_T^{IM} - P_T^{EX}$, equal to the length of the green line segment in [Figure 7.38 "Welfare Effects of an Export](#)

Tax: Large Country Case". If the tax were an ad valorem tax, then the tax rate would be given by

$$T = \frac{P_T^{IM}}{P_T^{EX}} - 1.$$

Table 7.17 "Welfare Effects of an Export Tax" provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the governments in the importing and exporting countries. The aggregate national welfare effects and the world welfare effects are also shown.

Table 7.17 Welfare Effects of an Export Tax

	Importing Country	Exporting Country
Consumer Surplus	- (A + B + C + D)	+ e
Producer Surplus	+ A	- (e + f + g + h)
Govt. Revenue	0	+ (c + g)
National Welfare	- (B + C + D)	+ c - (f + h)
World Welfare	- (B + D) - (f + h)	

Refer to Table 7.17 "Welfare Effects of an Export Tax" and Figure 7.38 "Welfare Effects of an Export Tax: Large Country Case" to see how the magnitudes of the changes are represented.

Export tax effects on the exporting country's consumers. Consumers of the product in the exporting country experience an increase in well-being as a result of the export tax. The decrease in their domestic price raises the amount of consumer surplus in the market.

Export tax effects on the exporting country's producers. Producers in the exporting country experience a decrease in well-being as a result of the tax. The decrease in the price of their product in their own market decreases producer surplus in the industry. The price decline also induces a decrease in output, a decrease in employment, and a decrease in profit, payments, or both to fixed costs.

Export tax effects on the exporting country's government. The government receives tax revenue as a result of the export tax. Who benefits from the revenue depends on how the government spends it. Typically, the revenue is simply included as part of

the general funds collected by the government from various sources. In this case, it is impossible to identify precisely who benefits. However, these funds help support many government spending programs, which presumably help either most people in the country, as is the case with public goods, or certain worthy groups. Thus someone within the country is the likely recipient of these benefits.

Export tax effects on the exporting country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers and producers. The net effect consists of three components: a positive terms of trade effect (c), a negative consumption distortion (f), and a negative production distortion (h).

Because there are both positive and negative elements, the net national welfare effect can be either positive or negative. The interesting result, however, is that it can be *positive*. This means that an export tax implemented by a large exporting country *may* raise national welfare.

Generally speaking, the following are true:

1. Whenever a large country implements a small export tax, it will raise national welfare.
2. If the tax is set too high, national welfare will fall.
3. There will be a positive optimal export tax that will maximize national welfare.

However, it is also important to note that not everyone's welfare rises when there is an increase in national welfare. Instead, there is a redistribution of income.

Producers of the product and recipients of government spending will benefit, but consumers will lose. A national welfare increase, then, means that the sum of the gains exceeds the sum of the losses across all individuals in the economy.

Economists generally argue that, in this case, compensation from winners to losers can potentially alleviate the redistribution problem.

Export tax effects on the importing country's consumers. Consumers of the product in the importing country suffer a reduction in well-being as a result of the export tax. The increase in the price of both imported goods and the domestic substitutes reduces the amount of consumer surplus in the market.

Export tax effects on the importing country's producers. Producers in the importing country experience an increase in well-being as a result of the export tax. The increase in the price of their product on the domestic market increases producer surplus in the industry. The price increase also induces an increase in the output of

existing firms (and perhaps the addition of new firms), an increase in employment, and an increase in profit, payments, or both to fixed costs.

Export tax effects on the importing country's government. There is no effect on the importing country's government revenue as a result of the exporter's tax.

Export tax effects on the importing country. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the government. The net effect consists of three components: a negative terms of trade effect (C), a negative production distortion (B), and a negative consumption distortion (D).

Since all three components are negative, the export tax must result in a reduction in national welfare for the importing country. However, it is important to note that a redistribution of income occurs—that is, some groups gain while others lose. In this case, the sum of the losses exceeds the sum of the gains.

Export tax effects on world welfare. The effect on world welfare is found by summing the national welfare effects on the importing and exporting countries. By noting that the terms of trade gain to the exporter is equal to the terms of trade loss to the importer, the world welfare effect reduces to four components: the importer's negative production distortion (B), the importer's negative consumption distortion (D), the exporter's negative consumption distortion (f), and the exporter's negative production distortion (h). Since each of these is negative, the world welfare effect of the export tax is negative. The sum of the losses in the world exceeds the sum of the gains. In other words, we can say that an export tax results in a reduction in world production and consumption efficiency.

KEY TAKEAWAYS

- An export tax raises consumer surplus and lowers producer surplus in the exporter market.
- An export tax lowers producer surplus in the export market and raises it in the import country market.
- National welfare may rise or fall when a large country implements an export tax.
- For any country that is large in an export product, there is a positive optimal export tax.
- National welfare in the importing country falls when a large exporting country implements an export tax.
- An export tax of any size will reduce world production and consumption efficiency and thus cause world welfare to fall.

EXERCISE

1. Suppose there are two large countries, the United States and China. Assume that both countries produce and consume clothing. The United States imports clothing from China. Consider the trade policy action listed along the top row of the table below. In the boxes, indicate the effect of the policy on the variables listed in the first column. Use a partial equilibrium, perfect competition model to determine the answers. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 7.18 EFFECTS OF AN EXPORT TAX

	Chinese Implementation of an Export Tax
U.S. Domestic Consumer Price	
U.S. Domestic Consumer Welfare	
U.S. Domestic Producer Welfare	
U.S. National Welfare	
Chinese Producer Welfare	
Chinese Consumer Welfare	
Chinese National Welfare	

Chapter 8

Domestic Policies and International Trade

Increasingly, at international forums where policymakers are discussing international trade issues, the topic of discussion is not what trade policies countries are using but rather what domestic policies are in place. The reason is that in our interconnected and globalized world, the domestic policies affecting energy, the environment, labor markets, health, and many other matters will affect not only what happens at home but also what, and how much, is traded and invested, and thus the outcomes for producers and consumers abroad. In short, domestic policies have international repercussions.

This chapter explores several simple domestic policies and investigates how these policies can affect trade flows with other countries. It also examines the welfare effects of these policies and concludes with a very important insight: that trade policies can be duplicated with a combination of several domestic policies. The implications of this notable insight are explored.

8.1 Chapter Overview

LEARNING OBJECTIVES

1. Distinguish trade policies from domestic policies.
2. Identify different types of domestic policies.
3. Learn the effects of simple domestic policies on small trading economies.

Policy analysis in international trade theory generally emphasizes the analysis of *trade* policies specifically. **Trade policy**¹ includes any policy that *directly* affects the flow of goods and services between countries, including import tariffs, import quotas, voluntary export restraints, export taxes, export subsidies, and so on. During the 1980s and 1990s, as trade barriers came down, especially between developed countries, more and more attention was brought to the effects of certain **domestic policy**² types, including their international effects.

For example, there is increasing concern in the United States about the environmental and labor policies of many U.S. trade partners. With regard to environmental policies, some have argued that more lenient environmental regulations in many less-developed countries give firms in those countries a competitive edge relative to firms operating in the United States. The same argument is used in regard to labor practices. Many U.S. industry representatives argue that low foreign wages, lenient occupational safety regulations, and in some cases the use of child labor or prison labor give some countries a competitive edge in international markets.

1. Any policy that directly affects the flow of goods and services between countries, such as import tariffs, import quotas, voluntary export restraints, export taxes, and export subsidies.
2. Any type of tax or subsidy policy or any type of government regulation that targets the domestic behavior of firms or consumers.

In general, for small countries, domestic policies *will* affect domestic prices, production levels, trade flows, and welfare but *will not* affect foreign prices, production levels, and welfare. This means that countries like the United States may not need to worry much about domestic practices in very small countries. However, when a country is large in international markets, domestic policies will affect prices, production levels, profits, and welfare, both domestically and internationally.

Types of Domestic Policies

In general, any type of domestic tax or subsidy policy, or any type of government regulation that affects the behavior of firms or consumers, can be classified as a

domestic policy. There are a wide variety of these policies, any of which can have an impact on international trade.

For example, income taxes are levied on wages and capital incomes of individuals. Profit taxes are levied on the profits of businesses. Sales taxes are generally levied as a percentage of retail sales. In the United States, these taxes are popular within individual states. Excise taxes are specific taxes on particular commodities such as gasoline, alcohol, or cigarettes.

Some domestic government policies take the form of quantity restrictions. An example is controls on the amount of pollutants that industries can emit. Also, in most countries there are restrictions on the production and sale of many drugs. The United States prohibits the use of recreational drugs like marijuana and cocaine, as well as pharmaceuticals that have not been approved by the U.S. Food and Drug Administration.

Governments also provide subsidies for many purposes. They disburse research and development (R&D) subsidies to high-technology industries and encourage R&D through their defense spending contracts. Governments also give out educational subsidies (grants) and subsidize student loans. In agriculture, governments often have elaborate programs designed to raise the incomes of farmers, including the use of price floors, subsidized loans, payments to encourage fallow acreage, and so on. Although many domestic policies are complex regulations, the analysis here will focus on simple domestic tax and subsidy policies applied either to production or to consumption. Many of the insights learned in this analysis, however, do carry over to more complex situations.

Domestic Policy versus Trade Policy Price Effects

One of the most important distinctions between domestic policies and trade policies is the effect on prices. When a trade policy, such as a tariff, is implemented, a price wedge is driven between the domestic price and the foreign price of the good. The domestic producers of the product will receive a higher price for the goods they sell, and domestic consumers will pay the same higher price for the goods they purchase.

3. The price received by producers, inclusive of any subsidies collected or taxes paid.
4. The price paid by consumers, inclusive of any subsidies collected or taxes paid.

In the case of domestic policies, a wedge is driven between domestic prices for the good. For example, if a domestic production subsidy is implemented by a small country, it will raise the price producers receive when they sell their good (we'll call this the **producer price**³), but it will not affect the price paid by domestic consumers when they purchase the good (we'll call this the **consumer price**⁴). The foreign price would remain equal to the consumer price in the domestic country.

Note that we can also call the consumer price the “market price” since this is the price that would appear on a price tag in the domestic market.

If a domestic consumption tax is implemented by a small country, it will raise the domestic consumer price of the good but will not affect the domestic producer price. The foreign price will remain equal to the producer price in this case.

In general, trade policies will always maintain the equality between domestic consumer and producer prices but will drive a wedge between domestic prices and foreign prices. Domestic policies (at least production and consumption taxes and subsidies), in contrast, will drive a wedge between domestic consumption and production prices.

Domestic Policies as a Basis for Trade

One of the first points made in this section is that a domestic policy can be the basis for trade. In other words, even if trade would not occur otherwise between countries, it is possible to show that the imposition of domestic taxes or subsidies can induce international trade, even if a country is small in international markets. Two examples are analyzed.

The first case considers a small country initially in free trade that, by chance, has no desire to export or import a particular commodity. The country then imposes a production subsidy. The subsidy encourages domestic production, but because the country is open to international trade, the domestic consumer price remains the same. Since the price paid by consumers remains the same, so does domestic demand. All the extra production, then, is exported to the rest of the world. Thus a domestic production subsidy can cause a commodity to be exported.

The second case considers the same initial conditions in which a small country in free trade has no desire to trade. In this case, the country implements a consumption tax. The tax raises the price paid by consumers in the domestic market, and this reduces domestic demand. However, because open competition remains with the rest of the world, the domestic producers’ price, and therefore domestic production, remains the same. The excess production over demand would now be exported to the rest of the world. Thus a domestic consumption tax can cause a commodity to be exported.

It would be straightforward to show that a production tax or a consumption subsidy (such as a rebate) could cause a country to import a good from the rest of the world.

Welfare Effects of Domestic Policies in Small Trading Economies

If a small country is importing or exporting a commodity initially, a domestic policy will affect the quantity imported or exported; the prices faced by consumers or producers; and the welfare of consumers, producers, the government, and the nation. We consider two examples in this section.

In the first case, we consider a production subsidy implemented by a small country that initially is importing the commodity from the rest of the world. The production subsidy stimulates domestic production by raising the producers' price but has no effect on the world price or the domestic consumers' price. Imports fall as domestic production rises.

Producers receive more per unit of output by the amount of the subsidy, thus producer surplus (or welfare) rises. Consumers face the same international price before and after the subsidy, thus their welfare is unchanged. The government must pay the unit subsidy for each unit produced by the domestic firms, and that represents a cost to the taxpayers in the country. The net national welfare effect of the production subsidy is a welfare loss represented by a production efficiency loss. Note, however, that the national welfare loss arises under an assumption that there are no domestic distortions or imperfections. If market imperfections are present, then a production subsidy can improve national welfare (see especially the infant industry argument in [Chapter 9 "Trade Policies with Market Imperfections and Distortions"](#), [Section 9.5 "The Infant Industry Argument and Dynamic Comparative Advantage"](#)).

In the second case, we consider a consumption tax implemented by a small country that initially is importing the commodity from the rest of the world. The consumption tax inhibits domestic consumption by raising the consumers' price but has no effect on the world price or the domestic producers' price. Imports fall as domestic consumption falls.

Consumers pay more for each unit of the good purchased, thus consumer surplus (or welfare) falls. Producers face the same international price before and after the tax, thus their welfare is unchanged. The government collects tax revenue for each unit sold in the domestic market, and that facilitates greater spending on public goods, thus benefitting the nation. The net national welfare effect of the consumption tax is a welfare loss represented by a consumption efficiency loss. Note again, however, that the national welfare loss arises under an assumption that there are no domestic distortions or imperfections. If market imperfections are present, then a consumption tax can improve national welfare.

Equivalency between Domestic and Trade Policies

Once the effects of simple domestic tax and subsidy policies are worked out, it is straightforward to show that a combination of domestic policies can duplicate a trade policy. For example, if a country imposes a specific production subsidy and a specific consumption tax on a product imported into the country and if the tax and subsidy rates are set equal, then the effects will be identical to a specific tariff on imports set at the same rate. If a country exports the product initially, then a production subsidy and consumption tax set at the same rates will be identical to an export subsidy set at the same level. Finally, a production tax coupled with a consumption subsidy (a rebate) imposed on a product that is initially exported and set at the same rate is equivalent to an export tax.

These results are especially important in light of recent movements in the direction of trade liberalization. As each new free trade agreement is reached, or as tariff barriers come down because of World Trade Organization (WTO) / General Agreement on Tariffs and Trade (GATT) negotiations, it seems reasonable to expect the expansion of international trade. Indeed, it is the effect that trade expansion will have on economic efficiency and growth that inspires these agreements in the first place. However, because trade policies are equivalent to a combination of domestic policies, it is possible to thwart the effects of trade liberalization by adjusting one's domestic policies.

Thus suppose a country negotiates and implements a free trade agreement with another country. As shown in our economic models, trade liberalization is likely to benefit some groups at the expense of others. Two main losses arise from trade liberalization. First, import-competing firms would lose out due to the increase in competition from foreign firms. Second, the government would lose tariff revenue.

Groups affiliated with import-competing industries are likely to be reluctant to support a free trade agreement. If these groups (trade associations, labor unions, etc.) are politically powerful, the domestic government may look for ways to reduce the harmful effects of trade liberalization by changing some of its domestic policies. An obvious way to do so would be to offer subsidies of some sort to the industries that are expected to be hurt by the agreement.

The other problem with trade liberalization is that it reduces government revenue. In this era where balanced government budgets are extremely difficult to maintain and where budget deficits are the norm, substantial reductions in government revenue are a serious source of concern. This means that many trade-liberalizing countries are likely to look for ways to mitigate the revenue shortfall. One obvious solution is to raise domestic taxes of some sort.

Although it is unlikely that a country's adjustments to its domestic policies would completely offset the effects of trade liberalization, it is conceivable that such adjustments would have some effect. Thus it is important for trade negotiators to be aware of the potential for domestic policy substitutions to assure that trade liberalizations have a real effect on trade between the countries.

The equivalency between trade and domestic policies may also be relevant to some of the trade disputes between the United States and Japan. Because of the large trade surpluses Japan had with the United States during the 1980s and 1990s, some people in the United States charged Japan with having excessive barriers to trade. Japan had noted, though, that its average tariff rates were roughly equivalent to tariffs charged by the United States and the EU. In the late 1980s, U.S. policymakers focused on Japan's domestic policies as the source of trade problems. In particular, the United States noted that Japan's distribution system and practices such as keiretsu (business groupings) may have been preventing U.S. firms' access to the Japanese market. This led to discussions known as the "Structural Impediments Initiative." Although this section does not suggest that such effects were indeed occurring, it does show that domestic policies can have an impact on trade flows between countries. In other words, it is conceivable that a country's domestic practices and policies could inhibit the inflow of goods into a country and act like tariffs or quotas on imports.

KEY TAKEAWAYS

- Domestic policies include all policies targeted at domestic production, consumption, or other activities. They include production and consumption taxes and subsidies as well as income sales, property taxes, and domestic regulations.
- In contrast, trade policies are targeted directly at imports and exports such as import tariffs and quotas and export taxes and subsidies.
- Production and consumption taxes and subsidies can stimulate imports or exports to occur. In other words, domestic policies can cause international trade.
- Domestic production and consumption taxes and subsidies will affect the level of international trade with the rest of the world.
- An import tariff applied on an imported product is equivalent in its economic effects to a combination of a domestic production subsidy and a domestic consumption tax of equal value applied on the same product.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term for the full price received by producers per unit of a good sold inclusive of any taxes or subsidies.
 - b. The term for the full price that consumers of a good pay to acquire a good inclusive of any taxes or subsidies.
 - c. The term for a government policy that directly affects trade between countries.
 - d. The term for a government policy that directly affects domestic economic activity.
 - e. Of *domestic policy* or *trade policy*, this describes an import quota.
 - f. Of *domestic policy* or *trade policy*, this describes a 5 percent state sales tax collected on all retail purchases.
 - g. Of *domestic policy* or *trade policy*, this describes a regulation on fuel efficiency standards on all automobiles sold in the United States.

8.2 Domestic Production Subsidies

LEARNING OBJECTIVES

1. Distinguish domestic production subsidies from export policies.
2. Describe the motivations for government use of production subsidies.

A **domestic production subsidy**⁵ is a payment made by a government to firms in a particular industry based on the level of output or production. The subsidy can be specified either as an ad valorem subsidy (a percentage of the value of production) or as a specific subsidy (a dollar payment per unit of output). A domestic production subsidy is different from an export subsidy. A production subsidy provides a payment based on all production regardless of where it is sold. An export subsidy, on the other hand, only offers a payment to the quantity or value that is actually exported. An export subsidy is classified as a trade policy, whereas a production subsidy is a domestic policy.

Domestic production subsidies are generally used for two main reasons. First, subsidies provide a way of raising the incomes of producers in a particular industry. This is in part why many countries apply production subsidies on agricultural commodities: it raises the incomes of farmers. The second reason to use production subsidies is to stimulate output of a particular good. This might be done because the product is assumed to be critical for national security. This argument is sometimes used to justify subsidies to agricultural goods, as well as steel, motor vehicles, the aerospace industry, and many other products. Countries might also wish to subsidize certain industries if it is believed that the industries are important in stimulating growth of the economy. This is the reason many companies receive research and development (R&D) subsidies. Although R&D subsidies are not strictly production subsidies, they can have similar effects.

We will analyze the international trade effects of a domestic production subsidy using a partial equilibrium analysis. We will assume that the market in question is perfectly competitive and that the country is “small.” We will also ignore any benefits the policy may generate, such as creating a more pleasing distribution of income or generating valuable external effects. Instead, we will focus entirely on the producer, consumer, and government revenue effects of each policy.

5. A payment made by a government to firms in a particular industry based on output or production levels.

Next, we consider the effects of a production subsidy under two separate scenarios. In the first case, the subsidy is implemented in a country that is not trading with

the rest of the world. This case is used to show how a domestic policy can cause international trade. The second case considers the price and welfare effects of a production subsidy implemented by a country that is initially importing the good from the rest of the world.

KEY TAKEAWAYS

- Domestic production subsidies are paid to firms for producing a product, whereas export subsidies are paid only to firms that export the product.
- The export subsidy is classified as a trade policy, whereas the production subsidy is a domestic policy.
- Production subsidies are used either to support the incomes within a sector or to stimulate production because it is believed that production will have a subsequent benefit.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term describing the type of payment made by a government to a firm for each unit of a good the firm produces.
 - b. The term describing the type of payment made by a government to a firm as a percentage of the value of a good the firm produces.
 - c. Of *domestic policy* or *trade policy*, this describes a production subsidy.
 - d. Of *domestic policy* or *trade policy*, this describes a specific export subsidy.

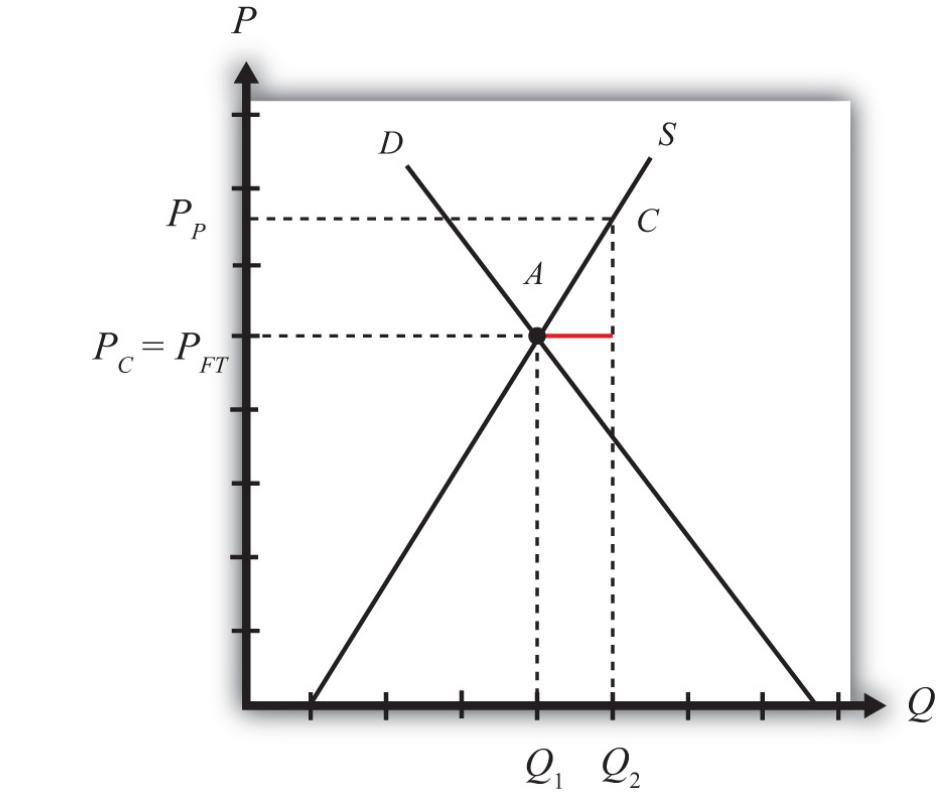
8.3 Production Subsidies as a Reason for Trade

LEARNING OBJECTIVE

1. Describe the price, quantity, and trade effects of a domestic production subsidy implemented by a small open economy.

This section will show how a production subsidy can *cause trade* for a small, perfectly competitive, open economy. The analysis indicates that domestic policies can be a cause of trade even in the absence of other reasons for trade. In other words, even if countries were identical with respect to their resource endowments, their technology, and their preferences and even if there were no economies of scale or imperfectly competitive markets, domestic policies could induce trade between countries.

Consider a small open economy with a perfectly competitive industry. Let the domestic market be represented by the supply and demand curves in [Figure 8.1 "Inducing Exports with a Domestic Production Subsidy"](#). Suppose initially that free trade is allowed with the rest of the world, but by coincidence (actually by assumption), let the free trade price be exactly equal to the autarky price for the good. This is shown as the price, P_{FT} . This implies that no imports or exports occur, even though there is free trade.

Figure 8.1 *Inducing Exports with a Domestic Production Subsidy*

Next, suppose that the government of this country offers a specific (per unit) production subsidy to the domestic firms. Let the subsidy rate be set at “ s .” This means the government will pay “ s ” dollars for every unit the domestic firm produces, regardless of where the product is sold.

The subsidy effectively raises the price that the producer receives for each unit of the good produced and sold. At the same time, the subsidy will not affect the domestic price that consumers pay. In other words, the subsidy will cause the price received by producers (the producer price) to rise above the price paid by consumers (the consumer price). The new producer price is labeled P_P in [Figure 8.1 "Inducing Exports with a Domestic Production Subsidy"](#), while the consumer price, P_C , remains equal to the free trade price. Thus $P_P = P_{FT} + s$ and $P_C = P_{FT}$. These price changes occur because these prices will allow domestic firms in the small country to maximize their profits in the face of free competition with firms in the rest of the world.

The subsidy will increase domestic production. At the market price P_{FT} , domestic firms were willing to supply to Q_1 . Once the producer price rises to P_P , domestic

supply will rise to Q_2 . Demand would remain the same, however, since the consumer price remains fixed. The difference between domestic supply and demand, $Q_2 - Q_1$, represents the level of exports to the rest of the world. Since exports did not exist prior to the subsidy, this is an example in which a domestic policy (a production subsidy) can cause trade (i.e., exports) to occur.

KEY TAKEAWAYS

- A production subsidy raises the price received by producers by the full amount of the subsidy when the country is open to international trade.
- A production subsidy has no effect on the price paid by consumers when the country is open to international trade.
- A production subsidy causes exports when implemented by a small country open to trade but not initially trading.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of exports or imports, the one that is likely to be increased as a result of a domestic production subsidy on that product.
 - b. Of *increase, decrease, or stay the same*, the effect on the producer price if a specific production subsidy is implemented by a country open to trade.
 - c. Of *increase, decrease, or stay the same*, the effect on the consumer price if a specific production subsidy is implemented by a country open to trade.

8.4 Production Subsidy Effects in a Small Importing Country

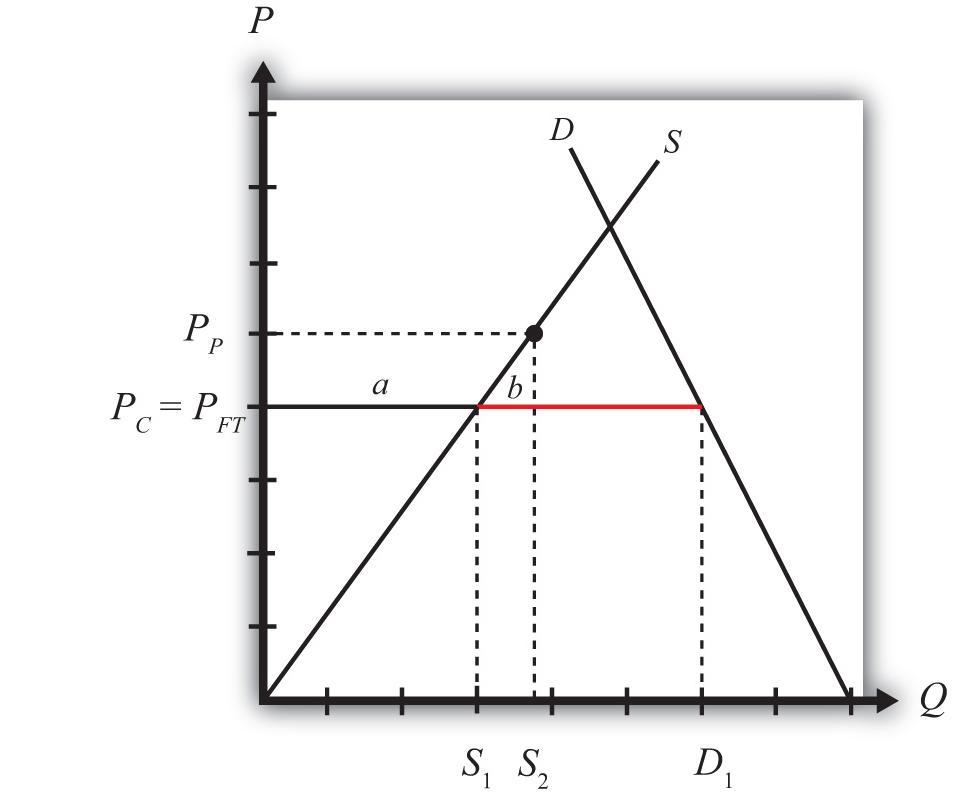
LEARNING OBJECTIVES

1. Identify the winners and losers when a small importing country implements a production subsidy.
2. Identify the national welfare effects when a small importing country implements a production subsidy.

Domestic policies can affect trade in an industry for a country that is either an exporter or an import-competitor initially. In this example, we consider the price, production, and welfare effects of a production subsidy when the subsidized product is initially imported into the country.

We depict this equilibrium in [Figure 8.2 "A Domestic Production Subsidy in a Small Importing Country"](#). The free trade price is given by P_{FT} . The domestic supply is S_1 , and domestic demand is D_1 , which determines imports in free trade as $D_1 - S_1$ (the length of the red line).

Figure 8.2 A Domestic Production Subsidy in a Small Importing Country



When a production subsidy “ s ” is imposed, the domestic producer price rises by the subsidy value to P_p . Because free trade is maintained and the importing country is small, the domestic consumer price remains at P_{FT} . Thus the effect of the subsidy in this case is to raise domestic supply from S_1 to S_2 while domestic demand remains at D_1 . As a result, imports fall from $(D_1 - S_1)$ to $(D_1 - S_2)$.

The welfare effects of the production subsidy are shown in [Table 8.1 "Static Welfare Effects of a Production Subsidy"](#). The letters in [Table 8.1 "Static Welfare Effects of a Production Subsidy"](#) refer to the areas labeled in [Figure 8.2 "A Domestic Production Subsidy in a Small Importing Country"](#).

Table 8.1 Static Welfare Effects of a Production Subsidy

	Importing Country
Consumer Surplus	0
Producer Surplus	+ a

	Importing Country
Govt. Revenue	- $(a + b)$
National Welfare	- b

Consumers are left unaffected by the subsidy since the domestic consumer price remains the same. Producers gain in terms of producer surplus. The subsidy causes the price producers receive to rise to P_P , which in turn stimulates an increase in output from S_1 to S_2 . The government, however, must pay the subsidy, and that means someone must pay higher taxes to fund it. The total amount of the subsidy payments is given by the product of $(P_P - P_{FT})$ in [Figure 8.2 "A Domestic Production Subsidy in a Small Importing Country"](#) (which corresponds to the subsidy rate) and the quantity produced, S_2 . Since the cost of the subsidy exceeds the benefits to producers, the net national welfare effect of the production subsidy is negative. Although one segment of the population benefits—namely, those connected with the import-competing industry—there remains a production efficiency loss, given by area b .

In the rest of the world, the small country assumption implies that this domestic policy (the production subsidy) would have no noticeable effects. Foreign prices would remain unchanged, and although their exports to this country would fall, these changes in trade volumes are too small to be noticed in the rest of the world. Thus the welfare effects on the rest of the world are said to be nonexistent, or zero.

KEY TAKEAWAYS

- A domestic production subsidy implemented in an import market by a small country will raise producer surplus for the import-competing firms, increase government expenditures and hence harm taxpayers, and leave consumers of the product unaffected.
- A domestic production subsidy implemented in an import market by a small country will create a net production efficiency loss and reduce national welfare.

EXERCISE

1. Consider the domestic policy action listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of the policy on the variables listed in the first column. Use a partial equilibrium model to determine the answers and assume that the shapes of the supply and demand curves are “normal.” Assume that the policy does not begin with, or result in, prohibitive policies. Also assume that the policy does not correct for market imperfections or distortions. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

For example, a production subsidy applied by a small country to an import-competing industry will have no effect on the domestic market price of the import good; therefore a 0 is placed in the first box of the table.

TABLE 8.2 EFFECTS OF A PRODUCTION SUBSIDY

	Production Subsidy to an Import Industry by a Small Country
Domestic Market Price	0
Domestic Industry Employment	
Domestic Consumer Welfare	

	Production Subsidy to an Import Industry by a Small Country
Domestic Producer Welfare	
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	
Foreign Consumer Welfare	
Foreign Producer Welfare	
Foreign National Welfare	

8.5 Domestic Consumption Taxes

LEARNING OBJECTIVES

1. Distinguish domestic consumption taxes from trade taxes.
2. Describe the motivations for government use of consumption taxes.

A **domestic consumption tax**⁶ is a tax collected by a government on sales of a particular product. The tax can be levied either as an ad valorem tax (a percentage of the value of the good) or as a specific tax (a charge per unit of the good sold). The domestic consumption tax is different from an import tariff or an export tax. The consumption tax is levied on all the goods sold in the domestic market regardless of where the goods are produced. An import tariff or export tax, on the other hand, is levied only on units of the goods actually imported or exported. An import tariff and an export tax are classified as trade policies, whereas the consumption tax is a domestic policy.

Domestic consumption taxes are often used as a source of government revenue. In the United States, the most common type of ad valorem consumption tax is the sales tax levied by state governments. The most common specific consumption taxes include gasoline, alcohol, and cigarette taxes. The latter two are sometimes referred to as “sin” taxes, since they are also designed to reduce consumption of potentially harmful substances. Thus sometimes consumption taxes are used to discourage certain types of consumption.

We will analyze the international trade effects of a domestic consumption tax using a partial equilibrium analysis. We will assume that the market in question is perfectly competitive and that the country is “small.” We will also ignore any benefits the policy may generate, such as creating a more pleasing distribution of income or generating valuable external effects. Instead, we will focus entirely on the producer, consumer, and government revenue effects of each policy.

Next, we consider the effects of a consumption tax under two separate scenarios. In the first case, the tax is implemented in a country that is not trading with the rest of the world. This case is used to show how a domestic policy can cause international trade. The second case considers the price and welfare effects of a consumption tax implemented by a country that is initially importing the good from the rest of the world.

6. A tax collected by a government on sales of a particular product.

KEY TAKEAWAYS

- Domestic consumption taxes are collected from consumers who purchase a product within the country, regardless of its country source, whereas tariffs and export taxes are collected only on the products that are imported or exported.
- An import tariff and an export tax are classified as trade policies, whereas the consumption tax is a domestic policy.
- Domestic consumption taxes are often collected to raise revenue for government expenditures.
- Domestic consumption taxes are sometimes used to discourage the consumption of some products.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term describing the type of payment received by a government for each unit of a good purchased by consumers.
 - b. The term describing the type of payment received by a government as a percentage of the value of a good purchased by consumers.
 - c. Of *domestic policy* or *trade policy*, this describes a consumption tax.
 - d. Of *domestic policy* or *trade policy*, this describes an export tax.

8.6 Consumption Taxes as a Reason for Trade

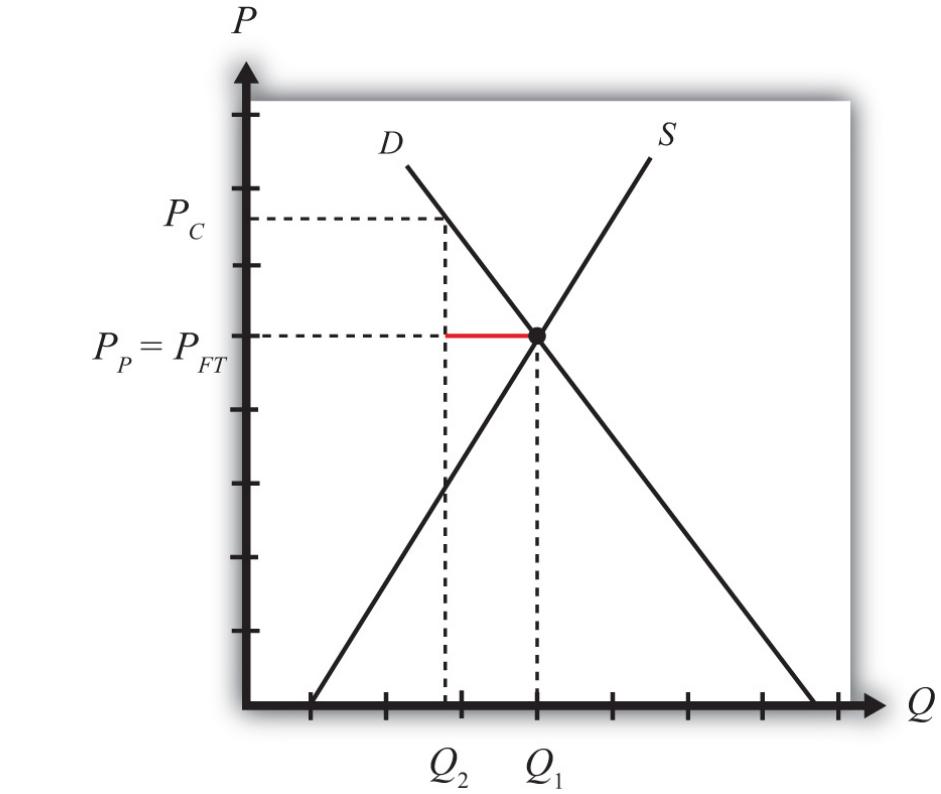
LEARNING OBJECTIVE

1. Describe the price, quantity, and trade effects of a domestic consumption tax implemented by a small open economy.

This section will show how a consumption tax can *cause trade* for a small, perfectly competitive, open economy. In other words, even if countries were identical with respect to their resource endowments, their technology, and their preferences and even if there were no economies of scale or imperfectly competitive markets, a purely domestic policy, such as a consumption tax, can induce trade between countries.

Consider a small open economy with a perfectly competitive industry. Let the domestic market be represented by the supply and demand curves in [Figure 8.3 "Inducing Exports with a Domestic Consumption Subsidy"](#). Suppose initially that free trade is allowed with the rest of the world, but by coincidence (actually by assumption), let the free trade price be exactly equal to the autarky price for the good. This is shown as the price, P_{FT} . At that price, both supply and demand equal Q_1 , and thus no imports or exports occur, even though there is free trade.

Figure 8.3 Inducing Exports with a Domestic Consumption Subsidy



Next, suppose that the government of this country imposes a specific (per unit) consumption tax on this product. Let the tax rate be set at “ t .” This means the government will collect “ t ” dollars for every unit of the good sold in the domestic market, regardless of whether the product is produced domestically or imported.

The tax will raise the domestic consumer price of the good by the full amount of the tax to P_C and reduce domestic demand to Q_2 . Domestic producers will not be affected by the consumption tax since continued competition in free trade with firms in the rest of the world will allow them to continue to charge the world price of P_{FT} . Note that in a closed economy, the producers would absorb some of the tax burden by lowering their price so as to maintain the profit maximum. However, being open to trade implies that the country can purchase as much as it likes at the world price. This means that the producer price P_P will remain equal to the free trade price P_{FT} , and the full burden of the tax falls on consumers. Thus $P_C = P_{FT} + t$ and $P_P = P_{FT}$.

Since the tax has no effect on the producer price but raises the consumer price, domestic demand falls to Q_2 while domestic supply remains at Q_1 . The difference, Q_1

- Q_2 (the length of the red line), represents the amount exported to the rest of the world. This implies that the consumption tax will induce exports of the good. Thus this is an example in which a domestic policy (a consumption tax) can cause trade (i.e., exports) to occur.

KEY TAKEAWAYS

- A consumption tax raises the price paid by consumers by the full amount of the tax when the country is open to international trade.
- A consumption tax has no effect on the price paid by producers when the country is open to international trade.
- A consumption tax causes exports when implemented by a small country open to trade but not initially trading.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the effect on the price consumers pay for a good when the government sets a domestic consumption tax in a freely trading economy.
 - b. Of *exports or imports*, the one that is likely to be increased as a result of a domestic consumption tax on that product.
 - c. Of *increase, decrease, or stay the same*, the effect on the price producers receive for a good when the government sets a domestic consumption tax in a freely trading economy.

8.7 Consumption Tax Effects in a Small Importing Country

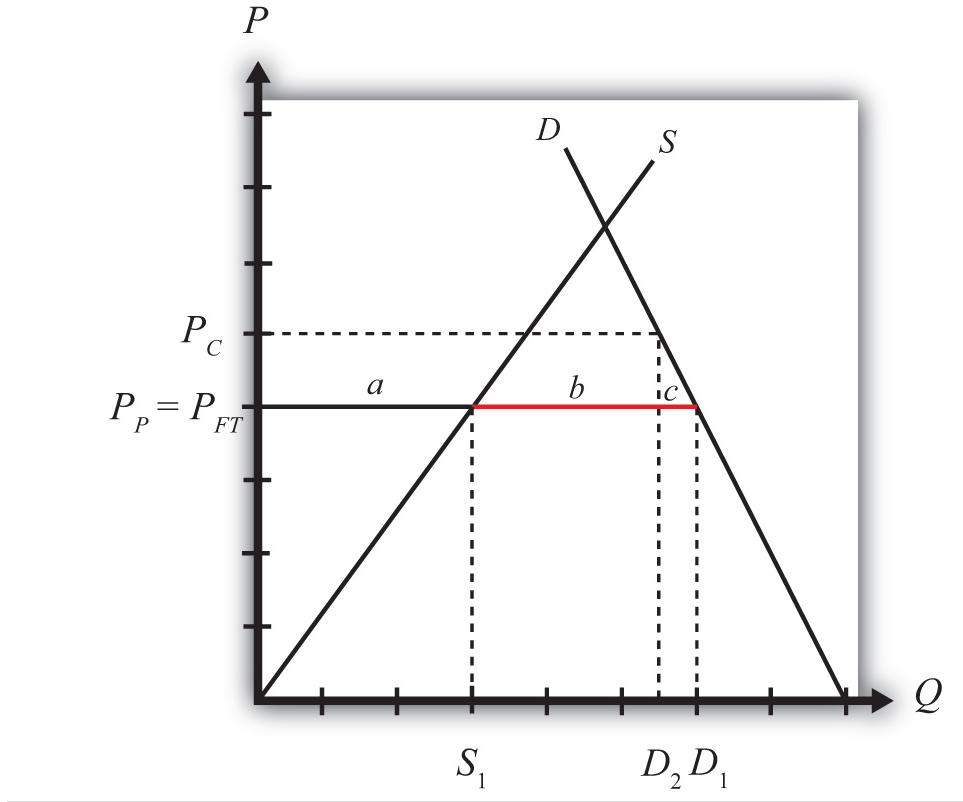
LEARNING OBJECTIVES

1. Identify the winners and losers when a small importing country implements a production subsidy.
2. Identify the national welfare effects when a small importing country implements a production subsidy.

Domestic policies can affect trade in an industry for a country that is either an exporter or an import-competitor initially. In this example, we consider the price, production, and welfare effects of a consumption tax when the taxed commodity is initially imported in the country.

We depict the initial equilibrium in [Figure 8.4 "A Domestic Consumption Tax in a Small Importing Country"](#). The free trade price is given by P_{FT} . The domestic supply is S_1 , and domestic demand is D_1 , which determines imports in free trade as $D_1 - S_1$ (the length of the red line).

Figure 8.4 A Domestic Consumption Tax in a Small Importing Country



When a specific consumption tax “ t ” is imposed, the consumer price will rise by the amount of the tax to P_C . The higher price paid by consumers will reduce their demand to D_2 . The producer price will remain at the free trade price indicated at $P_P = P_{FT}$, and hence domestic supply will remain at S_1 . The tax will reduce imports from $(D_1 - S_1)$ to $(D_2 - S_1)$.

The welfare effects of the consumption tax are shown in [Table 8.3 "Static Welfare Effects of a Consumption Tax"](#).

Table 8.3 Static Welfare Effects of a Consumption Tax

	Importing Country
Consumer Surplus	$- (a + b + c)$
Producer Surplus	0
Govt. Revenue	$+ (a + b)$
National Welfare	$- c$

Consumers suffer a loss in surplus because the price they pay rises by the amount of the consumption tax. Producers experience no change in surplus since the producer price (i.e., the price received by producers) remains at the free trade level. Note that even though imports fall, this decrease has no positive effect on producers in this situation. Finally, the government receives tax revenue from the consumption tax. The revenue is calculated as the tax, t (given by $P_C - P_P$), multiplied by the quantity consumed, D_2 .

Since the cost to consumers exceeds the benefits accruing to the government, the net national welfare effect of the consumption tax is negative. Although some segments of the population benefit, there remains a consumption efficiency loss, given by area c .

In the rest of the world, the small country assumption implies that this domestic policy (the consumption tax) would have no noticeable effects. Foreign prices would remain unchanged, and although their exports to this country would fall, these changes in trade volumes are too small to be noticed in the rest of the world. Thus the welfare effects on the rest of the world are said to be nonexistent, or zero.

KEY TAKEAWAYS

- A domestic consumption tax implemented in an import market by a small country will lower consumer surplus for domestic residents purchasing the product, increase government revenues and thereby benefit the recipients of subsequent government programs, and leave domestic producers of the product unaffected.
- A domestic consumption tax implemented in an import market by a small country will create a net consumption efficiency loss and reduce national welfare.

EXERCISE

1. Consider the domestic policy action listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of the policy on the variables listed in the first column. Use a partial equilibrium model to determine the answers and assume that the shapes of the supply and demand curves are “normal.” Assume that none of the policy does not begin with, or result in, prohibitive policies. Also assume that the policy does not correct for market imperfections or distortions. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 8.4 EFFECTS OF A CONSUMPTION TAX

	Consumption Tax on an Import Good by a Small Country
Domestic Market Price	
Domestic Industry Employment	
Domestic Consumer Welfare	
Domestic Producer Welfare	
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	

	Consumption Tax on an Import Good by a Small Country
Foreign Consumer Welfare	
Foreign Producer Welfare	
Foreign National Welfare	

8.8 Equivalence of an Import Tariff with a Domestic (Consumption Tax plus Production Subsidy)

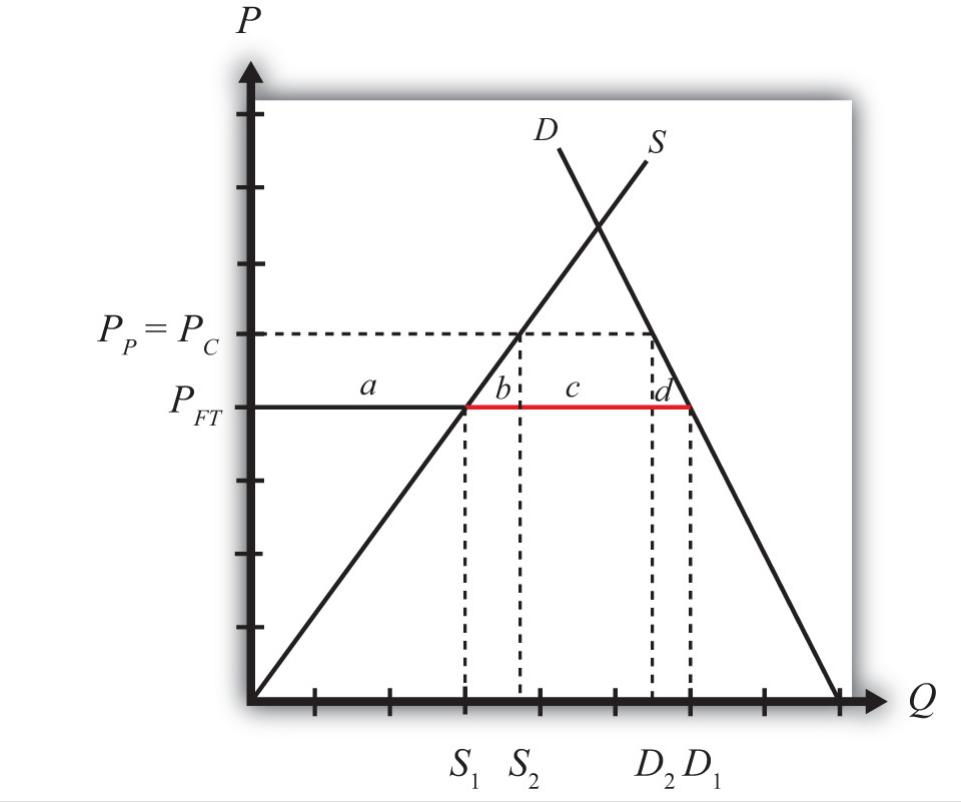
LEARNING OBJECTIVE

1. Learn that a combination of domestic policies can substitute for a trade policy.

We begin by demonstrating the effects of a consumption tax and a production subsidy applied simultaneously by a small importing country. Then we will show why the net effects are identical to an import tariff applied in the same setting and at the same rate.

In [Figure 8.5 "A Domestic Production Subsidy and Consumption Tax in a Small Importing Country"](#), the free trade price is given by P_{FT} . The domestic supply is S_1 , and domestic demand is D_1 , which determines imports in free trade as $D_1 - S_1$ (the red line).

Figure 8.5 A Domestic Production Subsidy and Consumption Tax in a Small Importing Country



When a specific consumption tax “ t ” is implemented, the consumer price increases by the amount of the tax to P_C . Because free trade is maintained, the producer’s price would remain at P_{FT} . The increase in the consumer price reduces domestic demand to D_2 .

When a specific production subsidy “ s ” is implemented, the producer price will rise by the amount of the tax to P_P , but it will not affect the consumer price. As long as the production subsidy and the consumption tax are set at the same value (i.e., $t = s$), which we will assume, the new producer price will equal the new consumer price (i.e., $P_C = P_P$).

The effect of the production subsidy and the consumption tax together is to lower imports from $D_1 - S_1$ to $D_2 - S_2$.

The combined welfare effects of the production subsidy and consumption tax are shown in [Table 8.5 "Static Welfare Effects of a Production Subsidy plus Consumption Tax"](#).

Table 8.5 Static Welfare Effects of a Production Subsidy plus Consumption Tax

	Importing Country
Consumer Surplus	- $(a + b + c + d)$
Producer Surplus	+ a
Net Govt. Revenue	+ c
Tax Revenue	+ $(a + b + c)$
Subsidy Cost	- $(a + b)$
National Welfare	- $(b + d)$

Consumers suffer a loss in surplus because the price they pay rises by the amount of the consumption tax.

Producers gain in terms of producer surplus. The production subsidy raises the price producers receive by the amount of the subsidy, which in turn stimulates an increase in output.

The government receives tax revenue from the consumption tax but must pay for the production subsidy. However, since the subsidy and tax rates are assumed to be identical and since consumption exceeds production (because the country is an importer of the product), the revenue inflow exceeds the outflow. Thus the net effect is a gain in revenue for the government.

In the end, the cost to consumers exceeds the sum of the benefits accruing to producers and the government; thus the net national welfare effect of the two policies is negative.

Notice that these effects are identical to the effects of a tariff applied by a small importing country if the tariff is set at the same rate as the production subsidy and the consumption tax. If a specific tariff, " t ," of the same size as the subsidy and tax were applied, the domestic price would rise to $P_T = P_{FT} + t$. Domestic producers, who are not charged the tariff, would experience an increase in their price to P_T . The consumer price would also rise to P_T . This means that the producer and consumer welfare effects would be identical to the case of a production subsidy and a consumption tax. The government would only collect a tax on the imported commodities, which implies tariff revenue given by (c) . This is exactly equal to the net revenue collected by the government from the production subsidy and consumption tax combined. The net national welfare losses to the economy in both

cases are represented by the sum of the production efficiency loss (*b*) and the consumption efficiency loss (*d*).

So What?

This equivalence is important because of what might happen after a country liberalizes trade. Many countries have been advised by economists to reduce their tariff barriers in order to enjoy the efficiency benefits that will come with open markets. However, any small country contemplating trade liberalization is likely to be faced with two dilemmas.

First, tariff reductions will quite likely reduce tariff revenue. For many developing countries today, tariff revenue makes up a substantial portion of the government's total revenue, sometimes as much as 20 percent to 30 percent. This is similar to the early days of currently developed countries. In the 1800s, tariff revenue made up as much as 50 percent of the U.S. federal government's revenue. In 1790, at the time of the founding of the nation, the U.S. government earned about 90 percent of its revenue from tariff collections. The main reason tariff revenue makes up such a large portion of a developing country's total government revenue is that tariffs are an administratively simple way to collect revenue. It is much easier than an income tax or profit tax, since those require careful accounting and monitoring. With tariffs, you simply need to park some guards at the ports and borders and collect money as goods come across.

The second problem caused by trade liberalization is that the tariff reductions will injure domestic firms and workers. Tariff reductions will cause domestic prices for imported goods to fall, reducing domestic production and producer surplus and possibly leading to layoffs of workers in the import-competing industries.

Trade-liberalizing countries might like to prevent some of these negative effects from occurring. This section then gives a possible solution. To make up for the lost tariff revenue, a country could simply implement a consumption tax. Consumption taxes are popular forms of taxation around the world. To mitigate the injury to its domestic firms, the country could implement production subsidies, which could forestall the negative impact caused by trade liberalization and could be paid for with extra revenue collected with the consumption tax.

This section demonstrates that if the consumption tax and production subsidy happened to be set on an imported product at equal values and at the same rate as the tariff reduction, then the two domestic policies would combine to fully duplicate the tariff's effects. In this case, trade liberalization would have no effect.

The General Agreement on Tariffs and Trade (GATT) and the World Trade Organization (WTO) agreements have always been cognizant of this particular possibility. The original text says that if after trade liberalization a country takes domestic actions nullifying the benefit that should accrue to the foreign export firms, then a country would be in violation of its GATT (or now WTO) commitments. In other words, it is a GATT/WTO violation to directly substitute domestic policies that duplicate the original effects of the tariff.

Nonetheless, even though a policy response like a production subsidy/consumption tax combination set only on trade liberalized products is unlikely, countries will still feel the effects of lost revenue and injury to import-competing producers. Thus countries will look for ways to compensate for the lost revenue and perhaps help out hard-hit industries.

This section shows that to the extent those responses affect imported products, they can somewhat offset the effects of trade liberalization. Thus it is well worth knowing that these equivalencies between domestic and trade policies are a possibility.

KEY TAKEAWAYS

- A domestic consumption tax on a product imported by a small country plus a domestic production subsidy set at the same rate as the tax has the same price and welfare effects as a tariff set at the same rate on the same imported product.
- The effects of trade liberalization could be offset with a domestic production subsidy and consumption tax combination on the imported good. However, these actions would be a WTO violation for WTO member countries.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The import policy equivalent to a combined domestic production subsidy and consumption tax applied on the same good at the same level.
 - b. Of *increase, decrease, or stay the same*, the effect on the domestic producer price with a combined domestic production subsidy and consumption tax applied on the same good at the same level.
 - c. Of *increase, decrease, or stay the same*, the effect on the domestic consumer price with a combined domestic production subsidy and consumption tax applied on the same good at the same level.
 - d. Of *increase, decrease, or stay the same*, the effect on the foreign price with a combined domestic production subsidy and consumption tax applied by a small country on the same good at the same level.

Chapter 9

Trade Policies with Market Imperfections and Distortions

Most models showing the advantages of international trade and the costs associated with protection assume that the world is perfectly competitive. The problem is that for a variety of reasons markets are usually not perfectly competitive, at least not completely so. Economists use the term “market imperfections” to describe situations that deviate from perfect competition. And when such deviations occur, interesting things happen.

For example, it is valid to say that in a world with market imperfections, free trade may not be the best policy to maximize national welfare; instead, some type of trade protection may be better. This chapter illustrates a series of examples with models that incorporate market imperfections to demonstrate this result. However, application of another theory in economics, the theory of the second best, and some other issues are shown to mitigate this result. In other words, even though trade policies can be used to raise a nation’s welfare, there may be a better way to achieve a superior result.

9.1 Chapter Overview

LEARNING OBJECTIVES

1. Understand that the presence of market imperfections or distortions in a trade model changes the potential outcomes of trade policies.
2. Learn the basic terminology used in discussing the theory of the second best.

Most of the models previously discussed incorporate a very standard economic assumption: namely, that markets are perfectly competitive. This was true in the Ricardian model, the Heckscher-Ohlin model, the specific factor model, and all the partial equilibrium analyses of trade and domestic policies using supply and demand curves in specific markets. The only deviation from perfect competition was in the discussion of economies-of-scale models and monopolistic competition. This is important because almost all the results concerning the effects of trade and trade policies presume that markets are perfectly competitive. But what if they're not?

Many critics of the economic conclusions about trade argue that the assumptions of perfect competition are unrealistic and that as a result standard trade theory misses some of the important impacts of trade found in the real world. There is much truth to this. By default, perfect competition models include many assumptions that are unrealistic. However, in defense, that is the nature of model building. Simplification is necessary to make the models tractable and solvable. If we were to try to create a model that included many or most of the complexities that we can imagine are present in real-world markets, we would no doubt quickly be overwhelmed with the model's intractability and might find it impossible to even identify an equilibrium solution. Indeed, in the real world, being in "equilibrium" might even be a rare occurrence.

Criticisms of economic theory along these lines, however, fail to recognize that economic analysis includes many attempts to incorporate market realities. Although it remains difficult to include many complexities simultaneously, it is possible to study them in a piecemeal way: one at a time.

The all-encompassing terms economists use to describe these complexities are market imperfections, or market failures, and market distortions. These cases are worthy of study because it is clear that markets rarely satisfy all the assumptions

made under perfect competition. These cases offer compelling arguments for protection, including the infant industry argument, the optimal tariff argument, strategic trade policy arguments, and arguments concerning national security.

Market imperfections or market distortions¹, generally, are any deviations from the assumptions of perfect competition. These include monopoly and oligopoly markets, production with increasing returns to scale, markets that do not clear, negative and positive externalities in production and consumption, and the presence of public goods.

When imperfections or distortions are present in a trade model, it is usually possible to identify a trade policy that can raise aggregate economic efficiency. In this chapter many cases are demonstrated in which trade policies improve national welfare. These welfare-improving policies, although detrimental to national welfare when used in a perfectly competitive setting, act to correct the imperfections or distortions present in the market. As long as the welfare impact of the correction exceeds the standard welfare loss associated with the trade policy, the policy will raise welfare.

Trade policies with market imperfections and distortions represent applications of the **theory of the second best**², formalized by Richard G. Lipsey and Kelvin Lancaster. See R. G. Lipsey and K. Lancaster, "The General Theory of the Second Best," *Review of Economic Studies* 24 (1956): 11–32. When imperfections or distortions are present in an international trade model, we describe the resulting equilibrium as second best. In this case, the standard policy prescriptions to maximize national welfare in a first-best or nondistorted economy will no longer hold true. Also, the implementation of what would be a detrimental policy in a first-best world can become a beneficial policy when implemented within a second-best world. For example, tariffs applied by a small country in the presence of domestic distortions can sometimes raise national welfare.

1. Any situation that deviates from the explicit or implicit assumptions of perfect competition.
2. Describes the class of models that consider policy implications in the presence of market imperfections and distortions.
3. The policy that raises welfare to the highest level possible; with market imperfections or distortions present, the policy that most directly corrects the distortion or imperfection.

In 1971, Jagdish Bhagwati presented a general theory of distortions in trade situations. See J. N. Bhagwati, "The Generalized Theory of Distortions and Welfare," in *Trade, Balance of Payments and Growth*, ed. J. N. Bhagwati, R. W. Jones, R. A. Mundell, and J. Vanek (Amsterdam: North-Holland Publishing Co., 1971). He characterized many of the distortions that can occur and considered which policies could be used to correct each distortion and raise national welfare. He considered not only trade policies but also domestic tax or subsidy policies. He showed that for most distortions, a trade policy is inferior (in terms of the extent to which it can raise national welfare) to other purely domestic policies. The most appropriate or **first-best policy**³, in general, would be the policy that most directly corrects the

distortion or imperfection present in the market. This chapter provides numerous examples of policy rankings and applications of this general rule.

In one case, a trade policy does prove to be first best. This is the case of a large import or export country in international markets. In this case, the first-best policy is the optimal tariff or the optimal export tax.

Thus the results of this section are somewhat schizophrenic. On the one hand, these models offer some of the most compelling arguments supporting protection. For example, one can easily use these models to justify protection when national defense is a concern, when unemployment may arise in a market, when trade causes environmental degradation, or when there are infant industries in a country. On the other hand, in almost all of these cases, a trade policy is not the most effective policy tool available to correct the problems caused by the distortion or imperfection.

Finally, when more complex markets are considered, as when there are multiple distortions or imperfections present simultaneously, our ability to identify welfare-improving policies rapidly diminishes. The theory of the second best states that correcting one distortion in the presence of many may not improve welfare even if the policy makes perfect sense within the partial equilibrium framework containing the one distortion. The reason is that correcting one distortion may have unintentional (and probably immeasurable) impacts in other sectors due to the presence of other distortions. For example, suppose a trade policy is implemented to correct an environmental problem. One might be able to measure the welfare costs of the trade policy and the environmental benefits that would accrue to society and conclude that the benefits exceed the costs. However, the trade policy will have an impact on prices and resource allocation, potentially spreading across numerous sectors. Suppose one other sector, adversely affected, generates positive spillover effects that act to raise well-being for some groups. Then it is conceivable that the loss of the positive spillover effects would more than outweigh the net benefit accruing to society due to the environmental improvement. This means that the well-intentioned and reasonably measured environmental trade policy could result in an unintentional welfare loss for the nation. The more complex is the economy and the more distortions and imperfections that are present, the more likely it is that we simply cannot know what the national effects of trade policies will be.

KEY TAKEAWAYS

- In the presence of market imperfections or distortions, free trade may no longer be the best policy, even for a small open economy.
- Although trade policies can sometimes generate national welfare improvements, trade policies are often second-best policies, meaning that there are other nontrade policies that are superior (called first-best policies).
- The first-best policy is the policy that targets and corrects the market imperfection as directly as possible.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term describing any assumption that represents a deviation from the standard assumptions of perfect competition.
 - b. The term describing a policy that most directly corrects the market imperfection or distortion in a market.
 - c. The name of the theory describing the class of models that consider policy implications in the presence of market imperfections and distortions.

9.2 Imperfections and Distortions Defined

LEARNING OBJECTIVES

1. Identify the various types of market imperfections and distortions.
2. Recognize that market imperfections and distortions are widespread in real-world markets.

Market imperfections and distortions, generally, are any deviations from the assumptions of perfect competition. Many of the assumptions in a perfectly competitive model are implicit rather than explicit—that is, they are not always stated.

Below are descriptions of many different types of imperfections and distortions. Perfect competition models assume the absence of these items.

Monopoly, Duopoly, and Oligopoly

Perhaps the most straightforward deviation from perfect competition occurs when there are a relatively small number of firms operating in an industry. At the extreme, one firm produces for the entire market, in which case the firm is referred to as a monopoly. A monopoly has the ability to affect both its output and the price that prevails on the market. A duopoly consists of two firms operating in a market. An oligopoly represents more than two firms in a market but less than the many, many firms assumed in a perfectly competitive market. The key distinction between an oligopoly and perfect competition is that oligopoly firms have some degree of influence over the price that prevails in the market.

Another key feature of these imperfectly competitive markets is that the firms within them make positive economic profits. The profits, however, are not sufficient to encourage entry of new firms into the market. In other words, free entry in response to profit is not possible. The typical method of justifying this is by assuming that there are relatively high fixed costs. High fixed costs, in turn, imply increasing returns to scale. Thus most monopoly and oligopoly models assume some form of imperfect competition.

Large Countries in International Trade

Surprisingly, “large” importing countries and “large” exporting countries have a market imperfection present. This imperfection is more easily understood if we use the synonymous terms for “largeness,” monopsony and monopoly power. Large importing countries are said to have “**monopsony power in trade**⁴,” while large exporting countries are said to have “**monopoly power in trade**⁵.” Let’s first consider monopoly power.

When a large exporting country implements a trade policy, it will affect the world market price for the good. That is the fundamental implication of largeness. For example, if a country imposes an export tax, the world market price will rise because the exporter will supply less. An export tax set optimally will cause an increase in national welfare due to the presence of a positive terms of trade effect. This effect is analogous to that of a monopolist operating in its own market. A monopolist can raise its profit (i.e., its firm’s welfare) by restricting supply to the market and raising the price it charges its consumers. In much the same way, a large exporting country can restrict its supply to international markets with an export tax, force the international price up, and create benefits for itself with the terms of trade gain. The term monopoly “power” is used because the country is not a pure monopoly in international markets. There may be other countries exporting the product as well. Nonetheless, because its exports are a sufficiently large share of the world market, the country can use its trade policy in a way that mimics the effects caused by a pure monopoly, albeit to a lesser degree. Hence the country is not a monopolist in the world market but has “monopoly power” instead.

Similarly, when a country is a large importer of a good, we say that it has “monopsony power.” A monopsony represents a case in which there is a single buyer in a market where there are many sellers. A monopsony raises its own welfare or utility by restricting its demand for the product and thereby forcing the sellers to lower their price. By buying fewer units at a lower price, the monopsony becomes better off. In much the same way, when a large importing country places a tariff on imports, the country’s demand for that product on world markets falls, which in turn lowers the world market price. An import tariff set optimally will raise national welfare due to the positive terms of trade effect. The effects in these two situations are analogous. We say that the country has monopsony “power” because the country may not be the only importer of the product in international markets, yet because of its large size, it has “power” like a pure monopsony.

Externalities

4. Another term to describe a large importing country—that is, a country whose policy actions can affect international prices.

5. Another term to describe a large exporting country—that is, a country whose policy actions can affect international prices.

6. Economic actions that have effects *external* to the market in which the action is taken.

Externalities⁶ are economic actions that have effects *external* to the market in which the action is taken. Externalities can arise from production processes

(production externalities) or from consumption activities (consumption externalities). The external effects can be beneficial to others (positive externalities) or detrimental to others (negative externalities). Typically, because the external effects impact someone other than the producer or consumers, the producer and the consumers do not take the effects into account when they make their production or consumption decisions. We shall consider each type in turn.

Positive Production Externalities

Positive production externalities occur when production has a beneficial effect in other markets in the economy. Most examples of positive production externalities incorporate some type of learning effect.

For example, manufacturing production is sometimes considered to have positive spillover effects, especially for countries that are not highly industrialized. By working in a factory, the production workers and managers all learn what it takes to operate the factory successfully. These skills develop and grow over time, a process sometimes referred to as *learning by doing*. The skills acquired by the workers, however, are likely to spill over to others in the rest of the economy. Why? Because workers will talk about their experiences with other family members and friends. Factory managers may teach others their skills at local vocational schools. Some workers will leave to take jobs at other factories, carrying with them the skills that they acquired at the first factory. In essence, learning spillovers are analogous to infectious diseases. Workers who acquire skills in one factory in turn will *infect* other workers they come into contact with and will spread the *skill disease* through the economy.

A similar story is told concerning research and development (R&D). When a firm does R&D, its researchers learn valuable things about production that in turn are transmitted through the rest of the economy and have positive impacts on other products or production processes.

Negative Production Externalities

Negative production externalities occur when production has a detrimental effect in other markets in the economy. The negative effects could be felt by other firms or by consumers. The most common example of negative production externalities involves pollution or other environmental effects.

When a factory emits smoke into the air, the pollution will reduce the well-being of all the individuals who must breathe the polluted air. The polluted air will also

likely require more frequent cleaning by businesses and households, raising the cost incurred by them.

Water pollution would have similar effects. A polluted river cannot be used for recreational swimming or at least reduces swimmers' pleasures as the pollution rises. The pollution can also eliminate species of flora and fauna and change the entire ecosystem.

Positive Consumption Externalities

Positive consumption externalities occur when consumption has a beneficial effect in other markets in the economy. Most examples of positive consumption externalities involve some type of aesthetic effect.

Thus when homeowners landscape their properties and plant beautiful gardens, it benefits not only themselves but also neighbors and passersby. In fact, an aesthetically pleasant neighborhood where yards are neatly kept and homes are well maintained would generally raise the property values of all houses in the neighborhood.

One could also argue that a healthy lifestyle has positive external effects on others by reducing societal costs. A healthier person would reduce the likelihood of expensive medical treatment and lower the cost of insurance premiums or the liability of the government in state-funded health care programs.

Negative Consumption Externalities

Negative production externalities occur when consumption has a detrimental effect in other markets in the economy. Most examples of negative consumption externalities involve some type of dangerous behavior.

Thus a mountain climber in a national park runs the risk of ending up in a precarious situation. Sometimes climbers become stranded due to storms or avalanches. This usually leads to expensive rescue efforts, the cost of which is generally borne by the government and hence the taxpayers.

A drunk driver places other drivers at increased risk. In the worst outcome, the drunk driver causes the death of another. A smoker may also put others at risk if secondhand smoke causes negative health effects. At the minimum, cigarette smoke surely bothers nonsmokers when smoking occurs in public enclosed areas.

Public Goods

Public goods⁷ have two defining characteristics: nonrivalry and nonexcludability. **Nonrivalry**⁸ means that the consumption or use of a good by one consumer does not diminish the usefulness of the good to another. **Nonexcludability**⁹ means that once the good is provided, it is exceedingly costly to exclude nonpaying customers from using it. The main problem posed by public goods is the difficulty of getting people to pay for them in a free market.

The classic example of a public good is a lighthouse perched on a rocky shoreline. The lighthouse sends a beacon of light outward for miles, warning every passing ship of the danger nearby. Since two ships passing are equally warned of the risk, the lighthouse is nonrival. Since it would be impossible to provide the lighthouse services only to those passing ships that paid for the service, the lighthouse is nonexcludable.

The other classic example of a public good is national security or national defense. The armed services provide security benefits to everyone who lives within the borders of a country. Also, once provided, it is difficult to exclude nonpayers.

Information has public good characteristics as well. Indeed, this is one reason for the slow start of electronic information services on the World Wide Web. Once information is placed on a Web site, it can be accessed and used by millions of consumers almost simultaneously. Thus it is nonrival. Also, it can be difficult, although not impossible, to exclude nonpaying customers from accessing the services.

7. Goods that are nonrival (the consumption or use of a good by one consumer does not diminish the usefulness of the good to another) and nonexcludable (once the good is provided, it is exceedingly costly to exclude nonpaying customers from using it).

8. A situation in which consumption or use of a good by one consumer does not diminish the usefulness of the good to another.

9. A situation in which once the good is provided, it is exceedingly costly to exclude nonpaying customers from using it.

Nonclearing Markets

A standard assumption in general equilibrium models is that markets always clear—that is, supply equals demand at the equilibrium. In actuality, however, markets do not always clear. When markets do not clear, for whatever reason, the market is distorted.

The most obvious case of a nonclearing market occurs when there is unemployment in the labor market. Unemployment could arise if there is price stickiness in the downward direction, as when firms are reluctant to lower their wages in the face of restricted demand. Alternatively, unemployment may arise because of costly adjustment when some industries expand while others contract. As described in the immobile factor model, many factors would not immediately find alternative employment after being laid off from a contracting industry. In the interim, the factors must search for alternative opportunities, may need to relocate to another

geographical location, or may need to be retrained. During this phase, the factors remain unemployed.

Imperfect Information

One key assumption often made in perfectly competitive models is that agents have perfect information. If some of the participants in the economy do not have full and complete information in order to make decisions, then the market is distorted.

For example, suppose entrepreneurs did not know that firms in an industry were making positive economic profits. Without this information, new firms would not open to force economic profit to zero in the industry. As such, imperfect information can create a distortion in the market.

Policy-Imposed Distortions

Another type of distortion occurs when government policies are set in markets that are perfectly competitive and exhibit no other distortions or imperfections. These were labeled policy-imposed distortions by Jagdish Bhagwati since they do not arise naturally but rather via legislation.

Thus suppose the government of a small country sets a trade policy, such as a tariff on imports. In this case, the equilibrium that arises with the tariff in place is a distorted equilibrium.

KEY TAKEAWAYS

- An implicit assumption of perfect competition models is that there are no market imperfections or distortions in place.
- Among some of the most common market imperfections are monopolies, oligopolies, large countries in trade, externalities, public goods, nonclearing markets, imperfect information, and government tax and subsidy policies.
- Externality effects can arise from production or consumption activities.
- Externalities can be positive or negative in their effects.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe the favorable effect that a production activity can have in another market.
 - b. The term used to describe the detrimental effect that a consumption activity can have on another person.
 - c. The two characteristics that identify “public goods.”
 - d. The term used to describe the type of distortion that occurs when governments implement taxes, subsidies, or regulations in otherwise perfectly competitive markets.
 - e. The type of power a large importing country is said to have.
 - f. The type of power a large exporting country is said to have.

9.3 The Theory of the Second Best

LEARNING OBJECTIVES

1. Understand the key features of the theory of the second best.
2. Distinguish between first-best and second-best equilibria.
3. Distinguish between first-best and second-best policies.

The theory of the second best was formalized by Richard Lipsey and Kelvin Lancaster in 1956. The primary focus of the theory is what happens when the optimum conditions are not satisfied in an economic model. Lipsey and Lancaster's results have important implications for the understanding of not only trade policies but also many other government policies.

In this section, we will provide an overview of the main results and indicate some of the implications for trade policy analysis. We will then consider various applications of the theory to international trade policy issues.

First of all, one must note that economic models consist of exercises in which a set of assumptions is used to deduce a series of logical conclusions. The solution of a model is referred to as an equilibrium. An equilibrium is typically described by explaining the conditions or relationships that must be satisfied in order for the equilibrium to be realized. These are called the equilibrium conditions. In economic models, these conditions arise from the maximizing behavior of producers and consumers. Thus the solution is also called an optimum.

For example, a standard perfectly competitive model includes the following equilibrium conditions: (1) the output price is equal to the marginal cost for each firm in an industry, (2) the ratio of prices between any two goods is equal to each consumer's marginal rate of substitution between the two goods, (3) the long-run profit of each firm is equal to zero, and (4) supply of all goods is equal to demand for all goods. In a general equilibrium model with many consumers, firms, industries, and markets, there will be numerous equilibrium conditions that must be satisfied simultaneously.

Lipsey and Lancaster's analysis asks the following simple question: What happens to the other optimal equilibrium conditions when one of the conditions cannot be satisfied for some reason? For example, what happens if one of the markets does

not clear—that is, supply does not equal demand in that one market? Would it still be appropriate for the firms to set the price equal to the marginal cost? Should consumers continue to set each price ratio equal to their marginal rate of substitution? Or would it be better if firms and consumers deviated from these conditions? Lipsey and Lancaster show that, generally, when one optimal equilibrium condition is not satisfied, for whatever reason, all the other equilibrium conditions will change. Thus if one market does not clear, it would no longer be optimal for firms to set the price equal to the marginal cost or for consumers to set the price ratio equal to the marginal rate of substitution.

First-Best versus Second-Best Equilibria

Consider a small perfectly competitive open economy that has no market imperfections or distortions, no externalities in production or consumption, and no public goods. This is an economy in which all resources are privately owned, the participants maximize their own well-being, firms maximize profit, and consumers maximize utility—always in the presence of perfect information. Markets always clear and there are no adjustment costs or unemployment of resources.

The optimal government policy in this case is laissez-faire. With respect to trade policies, the optimal policy is free trade. Any type of tax or subsidy implemented by the government under these circumstances can only reduce economic efficiency and national welfare. Thus with a laissez-faire policy, the resulting equilibrium would be called *first best*. It is useful to think of this market condition as *economic nirvana* since there is no conceivable way of increasing economic efficiency at a **first-best equilibrium¹⁰**.

Of course, the real world is unlikely to be so perfectly characterized. Instead, markets will likely have numerous distortions and imperfections. Some production and consumption activities have externality effects. Some goods have public good characteristics. Some markets have a small number of firms, each of which has some control over the price that prevails and makes positive economic profit. Governments invariably set taxes on consumption, profit, property and assets, and so on. Finally, information is rarely perfectly and costlessly available.

10. A market equilibrium that arises in the absence of any market imperfections or distortions; in other words, under the standard assumptions of perfect competition.

Now imagine again a small, open, perfectly competitive economy with no market imperfections or distortions. Suppose we introduce one distortion or imperfection into such an economy. The resulting equilibrium will now be less efficient from a national perspective than when the distortion was not present. In other words, the introduction of one distortion would reduce the optimal level of national welfare.

In terms of Lipsey and Lancaster's analysis, the introduction of the distortion into the system would sever one or more of the equilibrium conditions that must be satisfied to obtain economic nirvana. For example, suppose the imperfection that is introduced is the presence of a monopolistic firm in an industry. In this case, the firm's profit-maximizing equilibrium condition would be to set its price greater than the marginal cost rather than equal to the marginal cost as would be done by a profit-maximizing perfectly competitive firm. Since the economic optimum obtained in these circumstances would be less efficient than in economic nirvana, we would call this equilibrium a **second-best equilibrium**¹¹. Second-best equilibria arise whenever all the equilibrium conditions satisfying economic nirvana cannot occur simultaneously. In general, second-best equilibria arise whenever there are market imperfections or distortions present.

Welfare-Improving Policies in a Second-Best World

An economic rationale for government intervention in the private market arises whenever there are uncorrected market imperfections or distortions. In these circumstances, the economy is characterized by a second-best rather than a first-best equilibrium. In the best of cases, the government policy can correct the distortions completely and the economy would revert back to the state under economic nirvana. If the distortion is not corrected completely, then at least the new equilibrium conditions, altered by the presence of the distortion, can all be satisfied. In either case, an appropriate government policy can act to correct or reduce the detrimental effects of the market imperfection or distortion, raise economic efficiency, and improve national welfare.

It is for this reason that many types of trade policies can be shown to improve national welfare. Trade policies, chosen appropriate to the market circumstances, act to correct the imperfections or distortions. This remains true even though the trade policies themselves would act to reduce economic efficiency if applied starting from a state of economic nirvana. What happens is that the policy corrects the distortion or imperfection and thus raises national welfare by more than the loss in welfare arising from the application of the policy.

Many different types of policies can be applied, even for the same distortion or imperfection. Governments can apply taxes, subsidies, or quantitative restrictions. They can apply these to production, to consumption, or to factor usage. Sometimes they even apply two or more of these policies simultaneously in the same market. Trade policies, like tariffs or export taxes, are designed to directly affect the flow of goods and services between countries. Domestic policies, like production subsidies or consumption taxes, are directed at a particular activity that occurs within the country but is not targeted directly at trade flows.

11. A market equilibrium that arises in the presence of one or more market imperfections or distortions.

One prominent area of trade policy research focuses on identifying the optimal policy to be used in a particular second-best equilibrium situation. Invariably, this research has considered multiple policy options in any one situation and has attempted to rank order the potential policies in terms of their efficiency-enhancing capabilities. As with the ranking of equilibria described above, the ranking of policy options is also typically characterized using the first-best and second-best labels.

Thus the ideal or optimal policy choice in the presence of a particular market distortion or imperfection is referred to as a first-best policy. The first-best policy will raise national welfare, or enhance aggregate economic efficiency, to the greatest extent possible in a particular situation.

Many other policies can often be applied, some of which would improve welfare. If any such policy raises welfare to a lesser degree than a first-best policy, then it would be called a **second-best policy**¹². If there are many policy options that are inferior to the first-best policy, then it is common to refer to them all as second-best policies. Only if one can definitively rank three or more policy options would one ever refer to a third-best or fourth-best policy. Since these rankings are often difficult, third-best (and so on) policies are not commonly denoted.

Trade Policies in a Second-Best World

In a 1971 paper, Jagdish Bhagwati provided a framework for understanding the welfare implications of trade policies in the presence of market distortions. See J. N. Bhagwati, “The Generalized Theory of Distortions and Welfare,” in *Trade, Balance of Payments and Growth*, ed. J. N. Bhagwati, R. W. Jones, R. A. Mundell, and J. Vanek (Amsterdam: North-Holland Publishing Co., 1971). This framework applied the theory of the second best to much of the welfare analysis that had been done in international trade theory up until that point. Bhagwati demonstrated the result that trade policies can improve national welfare if they occur in the presence of a market distortion and if they act to correct the detrimental effects caused by the distortion. However, Bhagwati also showed that in almost all circumstances a trade policy will be a second-best rather than a first-best policy choice. The first-best policy would likely be a purely domestic policy targeted directly at the distortion in the market. One exception to this rule occurs when a country is “large” in international markets and thus can affect international prices with its domestic policies. In this case, as was shown with optimal tariffs, quotas, voluntary export restraints (VERs), and export taxes, a trade policy is the first-best policy.

12. A policy whose best effect is inferior to another policy.

Since Bhagwati’s paper, international trade policy analysis has advanced to include market imperfections such as monopolies, duopolies, and oligopolies. In many of

these cases, it has been shown that appropriately chosen trade policies can improve national welfare. The reason trade policies can improve welfare, of course, is that the presence of the market imperfection means that the economy begins at a second-best equilibrium. The trade policy, if properly targeted, can reduce the negative aggregate effects caused by the imperfection and thus raise national welfare.

Summary of the Theory of the Second Best

In summary, the theory of the second best provides the theoretical underpinning to explain many of the reasons that trade policy can be shown to be welfare enhancing for an economy. In most (if not all) of the cases in which a trade policy is shown to improve national welfare, the economy begins at an equilibrium that can be characterized as second best. Second-best equilibria arise whenever the market has distortions or imperfections present. In these cases, it is relatively straightforward to conceive of a trade policy that corrects the distortion or imperfection sufficiently to outweigh the detrimental effects of the policy itself. In other words, whenever market imperfections or distortions are present, it is always theoretically or conceptually possible to design a trade policy that would improve national welfare. As such, the theory of the second best provides a rationale for many different types of protection in an economy.

The main criticism suggested by the theory is that rarely is a trade policy the first-best policy choice to correct a market imperfection or distortion. Instead, a trade policy is second best. The first-best policy, generally, would be a purely domestic policy targeted directly at the market imperfection or distortion.

In the remaining sections of this chapter, we use the theory of the second best to explain many of the justifications commonly given for protection or for government intervention with some form of trade policy. In each case, we also discuss the likely first-best policies.

KEY TAKEAWAYS

- A first-best equilibrium occurs in a perfectly competitive market when no imperfections or distortions are present.
- A second-best equilibrium arises whenever a market includes one or more imperfections or distortions.
- A first-best policy is that policy that can improve national welfare to the greatest extent when beginning in a second-best equilibrium.
- A second-best policy is one whose best national welfare effect is inferior to a first-best policy when beginning in a second-best equilibrium.
- As a general rule of thumb, beginning in a second-best equilibrium, the first-best policy will be a policy that attacks the market imperfection or distortion as directly as possible.
- As a general rule of thumb, domestic policies are usually first-best policies, whereas trade policies are usually second-best policies.
- One exception to the previous rule of thumb is that a trade policy is the first-best policy choice to correct the imperfection of a large country in international markets.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe an equilibrium that arises in the presence of market imperfections and distortions.
 - b. The term used to describe a policy action that can raise economic efficiency to the greatest extent possible.
 - c. The names of the economists who first formalized the theory of the second best.
 - d. The term used to describe an equilibrium that arises in the absence of market imperfections and distortions.
 - e. The term used to describe a policy action whose best effect is inferior to another policy option.

9.4 Unemployment and Trade Policy

LEARNING OBJECTIVES

1. Understand that unemployment of workers in a labor market is a type of market imperfection since supply of labor does not equal demand.
2. Recognize that a trade policy can be used to correct for an unemployment imperfection.
3. Learn the first-best and second-best policy options to correct for an unemployment imperfection in an import market.

Consider a small perfectly competitive economy. Suppose this economy has a market imperfection in the form of relatively immobile factors of production across industries. We will imagine that the labor force develops sector-specific skills as the time of employment in an industry increases. Thus if a worker works in an industry—say, the textile industry—for a long period of time, her productivity in textile production rises relative to nontextile workers who might begin employment in the textile industry. Similarly, other workers become more productive in their own industries relative to a textile worker who might begin employment in another industry.

These assumptions imply that although workers might be free to move across sectors of the economy, they might not be easily or costlessly transferred. Workers in one industry, accustomed to being paid a wage proportional to their productivity, might be unwilling to accept a lower wage in another industry even though the lower wage would reflect their productivity in that industry. A worker's reluctance to transfer could lead to a long search time between jobs as the worker continues to look for an acceptable job at an acceptable wage.

During the search period, a variety of adjustment costs would be incurred by the unemployed worker and by the government. The worker would suffer the anxiety of searching for another job. His or her family would have to adjust to a reduced income, and previous savings accounts would be depleted. At the worst, assets such as cars or homes may be lost. The government would compensate for some of the reduced income by providing unemployment compensation. This compensation would be paid out of tax revenues and thus represents a cost to others in the economy.

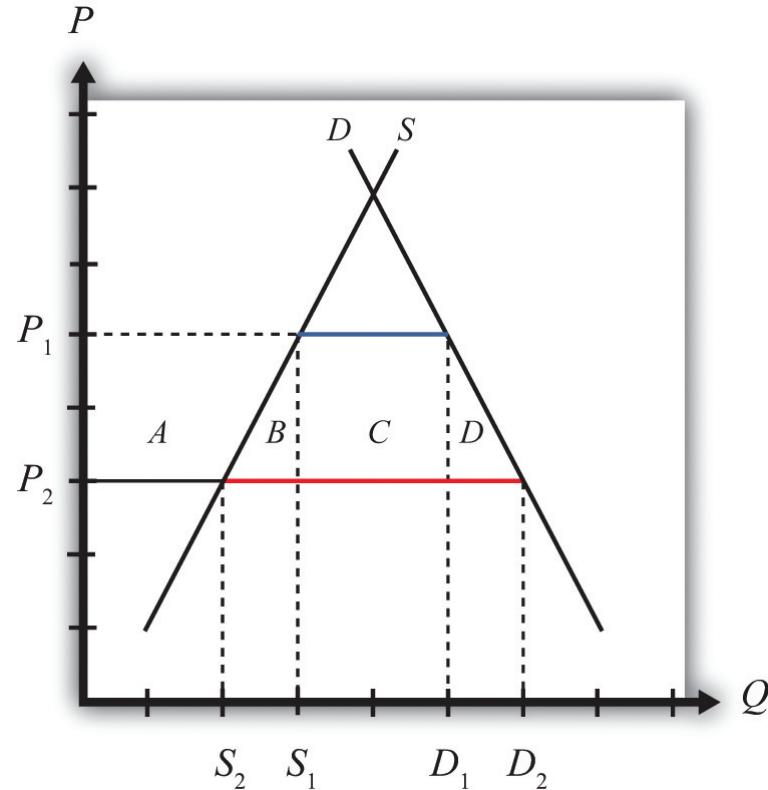
In some instances, the productivity of transferred workers could be raised by incurring training costs. These costs might be borne by the individual worker, as when the individual enrolls in a vocational training school. The costs might also be borne by an employer who hires initially low-productivity workers but trains them to raise their skills and productivity in the new industry.

In any case, the economy is assumed to have an unemployment imperfection that arises whenever resources must be transferred across industries. In every other respect, assume the economy is a small open economy with perfectly competitive markets and no other distortions or imperfections.

In the standard case of a small perfectly competitive economy, the optimal trade policy is free trade. Any tariff or quota on imports, although beneficial to the import-competing industry, will reduce aggregate efficiency—that is, the aggregate losses will exceed the aggregate benefits.

Imagine, however, that the economy initially has full employment of labor but that it has the unemployment imperfection described above. Suppose that initially the free trade price of textiles is given by P_1 in [Figure 9.1 "Unemployment in a Small Country Import Market"](#). At that price, demand is given by D_1 , supply by S_1 , and imports by $D_1 - S_1$ (the blue line segment).

Figure 9.1 *Unemployment in a Small Country Import Market*



Suppose that international market conditions suddenly change such that a surge of imports begins in the textile industry.

The surge can be represented by a reduction in the world price of the imported good from P_1 to P_2 . This would occur if there is an increase in total world supply of textiles of sufficient size to reduce the world price of the good. Since this importing country is assumed to be small, it must take the world price as given.

Domestic import-competing textile firms, to maintain profitability, would adjust to the lower free trade price by reducing output; supply would fall from S_1 to S_2 . The lower price would stimulate demand for the product, which would rise to D_2 . Thus imports would rise to $D_2 - S_2$ (the red line segment). The welfare effects of the lower world price are shown in Table 9.1 "Welfare Effects of a Lower Free Trade Price".

Table 9.1 Welfare Effects of a Lower Free Trade Price

	Importing Country
Consumer Surplus	+ (A + B + C + D)
Producer Surplus	- A
Unemployment Cost	- F
National Welfare	(B + C + D) - F

Consumers benefit from the lower free trade price. Producers lose in terms of a reduction in producer surplus. However, the unemployment imperfection implies that there is an additional cost that is hidden in this analysis. For domestic firms to reduce output requires them to reduce variable costs of production, which will include layoffs of workers. This means that the adjustment to the new free trade equilibrium will cause unemployment and its associated costs. We'll represent these unemployment or adjustment costs by the variable F . Note that these costs do not appear in [Figure 9.1 "Unemployment in a Small Country Import Market"](#).

The national welfare effects of the import surge depend on how high the unemployment costs (F) are compared to the aggregate benefits ($B + C + D$). Thus the national welfare effect could be positive or negative.

Effects of an Import Tariff

It is possible to eliminate the costs of unemployment by applying a tariff on imports of textiles. Suppose in response to the sudden drop in the free trade price, the government responds by implementing a tariff equal to $P_1 - P_2$. In this case, the domestic price would rise by the amount of the tariff. Instead of facing the new world price P_2 , the domestic country will face the original price P_1 . The tariff would eliminate the unemployment in the industry by keeping the domestic price at the original level. Domestic supply would remain at S_1 and employment would also remain at its original level.

However, implementing the tariff will also impose other costs on the economy. [Table 9.2 "Welfare Effects of an Import Tariff"](#) provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the government in the importing country. These effects are calculated relative to the economic situation *after* the surge of imports occurs. The aggregate national welfare effects are also shown.

Table 9.2 Welfare Effects of an Import Tariff

	Importing Country
Consumer Surplus	- (A + B + C + D)
Producer Surplus	+ A
Govt. Revenue	+ C
Unemployment Cost	+ F
National Welfare	F - (B + D)

Tariff effects on the importing country's consumers. Consumers of the product in the importing country suffer a reduction in well-being as a result of the tariff. The increase in the domestic price of both imported goods and the domestic substitutes reduces the amount of consumer surplus in the market. Refer to [Table 9.2 "Welfare Effects of an Import Tariff"](#) and [Figure 9.1 "Unemployment in a Small Country Import Market"](#) to see how the magnitude of the change in consumer surplus is represented.

Tariff effects on the importing country's producers. Producers in the importing country experience an increase in well-being as a result of the tariff. The increase in the price of their product on the domestic market increases producer surplus in the industry. Refer to [Table 9.2 "Welfare Effects of an Import Tariff"](#) and [Figure 9.1 "Unemployment in a Small Country Import Market"](#) to see how the magnitude of the change in producer surplus is represented.

Tariff effects on the importing country's government. The government receives tariff revenue as a result of the tariff. Who benefits from the revenue depends on how the government spends it. Typically, the revenue is simply included as part of the general funds collected by the government from various sources. In this case, it is impossible to identify precisely who benefits. However, these funds help support many government spending programs that presumably either help most people in the country, as is the case with public goods, or target certain worthy groups. Thus someone within the country is the likely recipient of these benefits. Refer to [Table 9.2 "Welfare Effects of an Import Tariff"](#) and [Figure 9.1 "Unemployment in a Small Country Import Market"](#) to see how the magnitude of the tariff revenue is represented.

Unemployment Costs

The tariff eliminates the unemployment or adjustment costs that would have been incurred in the absence of protection. Hence welfare rises by the amount F .

The aggregate welfare effect for the importing country is found by summing the gains and losses to consumers, producers, the government, and the potentially unemployed workers. The net effect consists of three components: a positive effect on workers who are saved from the negative effects of unemployment (F), a negative production distortion (B), and a negative consumption distortion (D).

Whether the country benefits from protection in the presence of an unemployment imperfection depends on how the cost of unemployment compares with the standard aggregate welfare cost of protection. If the aggregate costs of unemployment (F) that would arise in the absence of a tariff exceed the deadweight costs of the tariff (i.e., $B + D$), then national welfare would rise when the tariff is implemented. The tariff would eliminate the adjustment costs of unemployment while imposing other lower costs on consumers who would lose the benefit of lower prices.

With a more completely specified model, one could determine the optimal level of protection in these circumstances. It is not necessarily true that the optimal tariff will be the tariff that maintains the price at the original level. Instead, the optimal tariff will be achieved when the marginal cost of raising it further is just equal to the marginal benefit of the reduction in unemployment costs. This may be lower than the level set in the example above.

Objections to Protection

Of course, it is also conceivable that the aggregate costs of the tariff ($B + D$) exceed the aggregate adjustment costs (F) incurred by those who would become unemployed. In this case, the optimal tariff would remain zero and it would be best for the country to allow the adjustment to proceed. Thus the mere presence of unemployment is not sufficient evidence to justify the use of protection.

Also, even if protection is beneficial in the aggregate, it is important to remember that protection generates a redistribution of income. A tariff will force consumers to pay higher prices than they would have to pay in free trade. The extra costs to consumers are essentially being transferred to the firms and workers in the import-competing industry and to the government in the form of tariff revenue.

Finally, one could object to protection by noting that the benefit of protection—that is, eliminating unemployment—represents the permanent avoidance of temporary costs. If free trade were maintained in the face of the import surge, unemployment and its associated costs would be incurred, but these costs are likely to be temporary. Eventually workers will find alternative employment opportunities in other industries and the adjustment costs will dissipate. However, the benefits of free trade in the form of lower prices for consumers would be permanent benefits. Lower prices would presumably prevail period after period into the future. This means that even if the one-period benefits of eliminating unemployment exceed the one-period costs of protection, this may not hold if evaluated over multiple periods.

First-Best versus Second-Best Policies

Another objection to the use of a tariff to eliminate the cost of unemployment is that a tariff will be a second-best policy to correct the unemployment imperfection. The first-best policy would be a policy targeted more directly at the source of the market imperfection—in this case, the unemployment. Many such policies would be superior to a tariff. One easy-to-analyze policy is a production subsidy. A production subsidy means that the government would make payments, say, per unit of output produced by the domestic firms.

Begin with the same surge of imports described in [Figure 9.1 "Unemployment in a Small Country Import Market"](#) in the import market and with the same welfare costs and benefits. This time, however, suppose that the government offers a production subsidy sufficient to raise output in the domestic industry back to the original level. Recall that a production subsidy will raise the producer's price by the amount of the subsidy for a small country and will maintain the consumer price at its original level. A specific production subsidy “ s ” set equal to the difference $P_1 - P_2$ would cause the producer price to rise to P_1 while the consumer price would remain at P_2 . The higher producer price will induce domestic firms to raise their supply back to the original level of S_1 , but the constant consumer price will keep domestic demand at D_2 .

[Table 9.3 "Welfare Effects of a Production Subsidy"](#) provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the government in the importing country as a result of the production subsidy. These effects are calculated relative to the economic situation *after* the surge of imports occurs. The aggregate national welfare effects are also shown.

Table 9.3 Welfare Effects of a Production Subsidy

	Importing Country
Consumer Surplus	0
Producer Surplus	+ A
Govt. Revenue	- (A + B)
Unemployment Cost	+ F
National Welfare	$F - B$

Production subsidy effects on the importing country's consumers. Consumers of the product in the importing country are unaffected by the subsidy since there is no change in the domestic price of the good.

Production subsidy effects on the importing country's producers. Producers in the importing country experience an increase in well-being as a result of the tariff. Although they receive the same free trade price in the market as before, they now also receive the per-unit subsidy payment from the government. That means that their surplus is measured off of the original supply curve. Refer to [Table 9.3 "Welfare Effects of a Production Subsidy"](#) and [Figure 9.1 "Unemployment in a Small Country Import Market"](#) to see how the magnitude of the change in producer surplus is represented.

Production subsidy effects on the importing country's government. The government must pay the per-unit production subsidy. The per-unit subsidy rate is given as the price difference ($P_1 - P_2$), while the quantity of domestic production is given by S_1 . The product of these two terms gives the value of the subsidy payments made by the government. Who loses from the subsidy payments depends on where the tax revenue is collected. Generally, it is impossible to identify precisely which taxpayers lose. Refer to [Table 9.3 "Welfare Effects of a Production Subsidy"](#) and [Figure 9.1 "Unemployment in a Small Country Import Market"](#) to see how the magnitude of the subsidy payments is represented.

Unemployment Costs

The subsidy eliminates the unemployment or adjustment costs that would have been incurred in the absence of the subsidy. Hence welfare rises by the amount F .

The aggregate welfare effect for the importing country is found by summing the gains and losses to consumers, producers, the government, and the potentially

unemployed workers. The net effect consists of two components: a positive effect on workers who are saved from the negative effects of unemployment (F) and a negative production distortion (B).

Whether the country benefits from a production subsidy in the presence of an unemployment imperfection depends on how the cost of unemployment compares with the standard aggregate welfare cost of protection. If the aggregate costs of unemployment (F) that would arise in the absence of a tariff exceed the production efficiency losses of the subsidy (i.e., B), then national welfare would rise when the production subsidy is implemented. The production subsidy would eliminate the adjustment costs of unemployment but would cost the taxpayer extra money to finance the subsidy.

However, the key difference is the comparison of the production subsidy with the import tariff. Both policy actions could generate an improvement in national welfare, but the production subsidy would raise national welfare by more than the import tariff. In [Figure 9.1 "Unemployment in a Small Country Import Market"](#), it can be seen that $F - B > F - B - D$. For this reason, we might refer to the production subsidy as a first-best policy, while the import tariff is second best.

The production subsidy is superior because it corrects the imperfection more directly. By targeting production, the production subsidy creates a production distortion (B) but eliminates an unemployment imperfection. The tariff, on the other hand, creates a production *and* consumption distortion ($B + D$) to eliminate the same unemployment imperfection. Generally, it is preferable to introduce as few other distortions as possible in designing a policy to correct another.

This example shows how a production subsidy is superior to a tariff. However, in the case of an unemployment imperfection, there are likely to be policies superior to the production subsidy. It would seem that some policies would target the imperfection even more directly.

For example, the government could use a labor employment subsidy if the primary problem were the potential unemployment of labor. In this case, the government would make a payment to firms for each worker hired. If set at the correct level, the subsidy could eliminate the negative effects caused by unemployment. However, since firms would remain free to substitute labor for other inputs, industry production levels might not be the same as with a production subsidy. Firms' freedom to adjust output could further reduce the cost of the additional distortion.

A labor employment subsidy, however, would not solve the problem of long-term adjustment. As mentioned, the cost associated with unemployment is likely to be temporary, while the cost of eliminating the unemployment with a subsidy would require a permanent taxpayer cost. Thus an even more superior policy would probably be one that is targeted even more directly at the source of the problem. Recall that the problem is in the adjustment process. Superior policies might be those that facilitate the adjustment of labor resources across industries.

In a sense, this is the purpose behind policies like trade adjustment assistance (TAA). TAA was originally implemented in the 1962 U.S. Trade Act. It provides for the extension of unemployment compensation, loans, and grants for technical retraining and other types of support programs for workers who are displaced as a result of trade liberalization. If TAA is designed and implemented in a cost-efficient manner, it could be first among the contenders for a first-best policy to correct an unemployment imperfection.

KEY TAKEAWAYS

- An import tariff that reduces unemployment costs sufficiently can raise national welfare, even for a small importing country.
- An import tariff is a second-best policy to correct for an unemployment imperfection after an import surge.
- A production subsidy is superior to an import tariff as a policy to correct for an unemployment imperfection after an import surge.
- A production subsidy might be classified as first best in this situation, except that even more targeted policies, like worker retraining, could be superior.
- In the presence of an unemployment imperfection after an import surge, a domestic policy is first best, while the best trade policy is second best.

EXERCISES

- Consider the following imperfect market situations in the table below. From the following list of policy options, identify all types of trade policies *and* all types of domestic policies that could potentially raise national welfare in the presence of each imperfection. Consider only the partial equilibrium effects of each policy.

Options: An import tariff, an import quota, a voluntary export restraint (VER), an export tax, an export subsidy, a production tax, a production subsidy, a consumption tax, and a consumption subsidy.

TABLE 9.4 WELFARE IMPROVING POLICIES

	Trade Policy	Domestic Policy
1. Unemployment in a small import-competing industry suffering from a surge of imports		
2. A small country in which an export decline causes unemployment		

- Consider the policy actions listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of each policy on the variables listed in the first column. Use a partial equilibrium model to determine the answers and assume that the shapes of the supply and demand curves are “normal.” Assume that none of the policies begin with, or result in, prohibitive policies. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 9.5 EFFECTS OF POLICIES TO ALLEVIATE UNEMPLOYMENT

	Import Tariff by a Small Country with Unemployment	Production Subsidy by a Small Country with Unemployment
Domestic Consumer Price		
Domestic Producer Price		
Domestic Industry Employment		
Unemployment Welfare Effect		
Domestic Consumer Welfare		
Domestic Producer Welfare		
Domestic Government Revenue		
Domestic National Welfare		

9.5 The Infant Industry Argument and Dynamic Comparative Advantage

LEARNING OBJECTIVES

1. Learn that the infant industry argument presumes a market imperfection—the presence of a positive production externality.
2. Recognize that a trade policy can be used to correct for an infant industry production externality imperfection.
3. Learn the first-best and second-best policy options to correct for an infant industry production externality imperfection.
4. Learn the practical implementation problems that can arise when governments attempt to apply infant industry protection.

One of the most notable arguments for protection is known as the **infant industry**¹³ argument. The argument claims that protection is warranted for small new firms, especially in less-developed countries. New firms have little chance of competing head-to-head with the established firms located in the developed countries. Developed country firms have been in business longer and over time have been able to improve their efficiency in production. They have better information and knowledge about the production process, about market characteristics, about their own labor market, and so on. As a result, they are able to offer their product at a lower price in international markets and still remain profitable.

A firm producing a similar product in a less-developed country (LDC), on the other hand, would not have the same production technology available to it. Its workers and management would lack the experience and knowledge of its developed country rivals and thus would most likely produce the product less efficiently. If forced to compete directly with the firms in the developed countries, the LDC firms would be unable to produce profitably and thus could not remain in business.

Protection of these LDC firms, perhaps in the form of an import tariff, would raise the domestic price of the product and reduce imports from the rest of the world. If prices are raised sufficiently, the domestic firms would be able to cover their higher production costs and remain in business. Over time, these LDC firms would gain production and management experience that would lower their production costs. Essentially, the firms would follow the same path that the developed country firms had followed to realize their own production efficiency improvements. Protection, then, allows an infant industry time to “grow up.”

13. An industry, most often in a developing country, that cannot compete in international markets in free trade but that, if given time to learn and develop, could be world-class efficient.

Furthermore, since the LDC firms would improve their productive efficiency over time, the protective tariffs could be gradually reduced until eventually, when the tariffs are eliminated, they would compete on an equal footing with the developed country firms.

Many people have argued that this was precisely the industrial development strategy that was pursued by countries like the United States and Germany during their rapid industrial development before the turn of the twentieth century. Both the United States and Germany had high tariffs during their industrial revolution periods. These tariffs helped protect fledgling industries from competition with more-efficient firms in Britain and may have been the necessary requirement to stimulate economic growth.

One counterargument to this theory is that by protecting infant industries, countries are not allocating resources in the short run on the basis of comparative advantage. The Ricardian and Heckscher-Ohlin models of trade show that resources will be allocated most efficiently if countries produce goods whose before-trade prices are lower than those in the rest of the world. This implies that the United States and Germany should have simply imported the cheaper industrial goods from Britain and shifted their own resources to other goods in which they had a comparative advantage if they wished to maximize economic efficiency.

The reason for the discrepancy in policy prescriptions can easily be seen by noting the difference between static comparative advantage and dynamic comparative advantage. The traditional Ricardian theory of comparative advantage identifies the most efficient allocation of resources at one point in time. In this sense, it is a *static* theory. The policy prescription is based on a snapshot in time.

On the other hand, the infant industry argument is based on a dynamic theory of comparative advantage. In this theory, one asks what is best for a country (i.e., what is most efficient) in the long run. The most efficient long-run strategy may well be different from what is best initially. Here's why.

The problem faced by many LDCs is that their static comparative advantage goods, in most instances, happen to be agricultural commodities and natural resources. Reliance on production of these two types of goods can be problematic for LDCs. First of all, the prices of agricultural commodities and natural resources have historically been extremely volatile. In some years prices are very high, and in other years the prices are very low. If a country allocates many of its resources to production of goods with volatile prices, then the gross domestic product (GDP) will fluctuate along with the prices. Some years will be very good, and others will be very bad. Although a wealthier country may be able to smooth income by

effectively using insurance programs, a poor country might face severe problems, perhaps as severe as famine, in years when the prices of their comparative advantage goods are depressed.

In addition, many people argue that the management and organizational skills necessary to produce agricultural goods and natural resources are not the same as the skills and knowledge needed to build an industrial economy. If true, then concentrating production in one's static comparative advantage goods would prevent the development of an industrial economy. Thus one of the reasons for protecting an infant industry is to stimulate the learning effects that will improve productive efficiency. Furthermore, these learning effects might spill over into the rest of the economy as managers and workers open new businesses or move to other industries in the economy. To the extent that there are positive spillovers or externalities in production, firms are unlikely to take account of these in their original decisions. Thus, if left alone, firms might produce too little of these types of goods and economic development would proceed less rapidly, if at all.

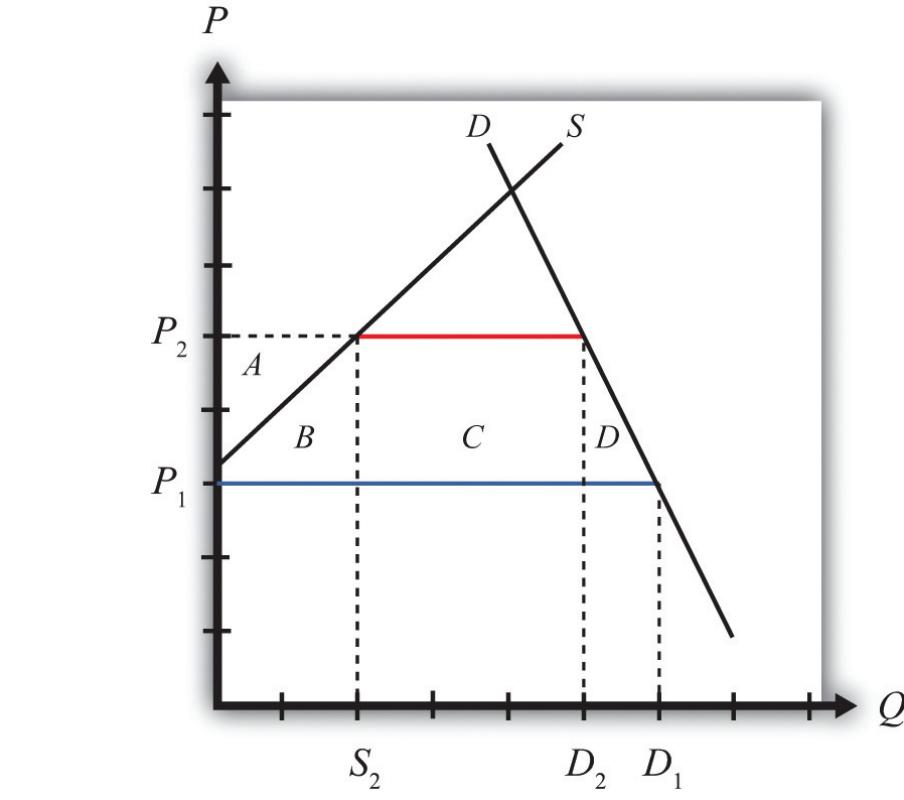
The solution suggested by the infant industry argument is to protect the domestic industries from foreign competition in order to generate positive learning and spillover effects. Protection would stimulate domestic production and encourage more of these positive effects. As efficiency improves and other industries develop, economic growth is stimulated. Thus by protecting infant industries a government might facilitate more rapid economic growth and a much faster improvement in the country's standard of living relative to specialization in the country's static comparative advantage goods.

An Analytical Example

Consider the market for a manufactured good such as textiles in a small, less-developed country.

Suppose that the supply and demand curves in the country are as shown in [Figure 9.2 "An Infant Industry in a Small Importing Country"](#). Suppose initially free trade prevails and the world price of the good is P_1 . At that price, consumers would demand D_1 , but the domestic supply curve is too high to warrant any production. This is the case, then, where domestic producers simply could not produce the product cheaply enough to compete with firms in the rest of the world. Thus the free trade level of imports would be given by the blue line segment, which is equal to domestic demand, D_1 .

Figure 9.2 An Infant Industry in a Small Importing Country



Suppose that the infant industry argument is used to justify protection for this currently nonexistent domestic industry. Let a specific tariff be implemented that raises the domestic price to P_2 . In this case, the tariff would equal the difference between P_2 and P_1 —that is, $t = P_2 - P_1$. Notice that the increase in domestic price is sufficient to stimulate domestic production of S_2 . Demand would fall to D_2 and imports would fall to $D_2 - S_2$ (the red line segment).

The static (i.e., one-period) welfare effects of the import tariff are shown in [Table 9.6 "Static Welfare Effects of a Tariff"](#).

Table 9.6 Static Welfare Effects of a Tariff

	Importing Country
Consumer Surplus	- (A + B + C + D)
Producer Surplus	+ A
Govt. Revenue	+ C

	Importing Country
National Welfare	- B - D

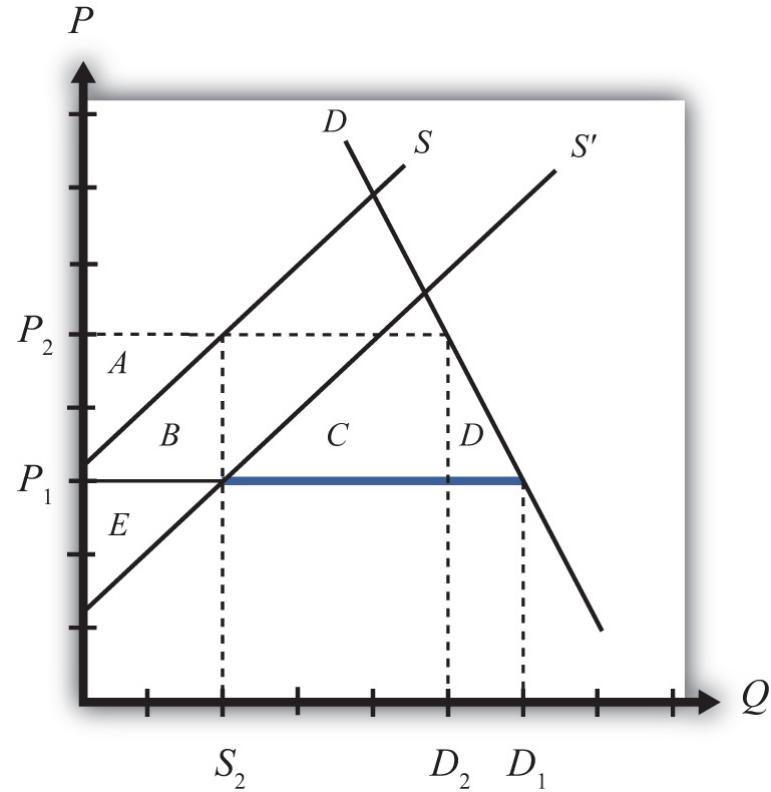
Consumers of textiles are harmed because of the higher domestic price of the good. Producers gain in terms of producer surplus. In addition, employment is created in an industry that did not even exist before the tariff. Finally, the government earns tariff revenue, which benefits some other segment of the population.

The net national welfare effect of the import tariff is negative. Although some segments of the population benefit, two deadweight losses to the economy remain. Area *B* represents a production efficiency loss, while area *D* represents a consumption efficiency loss.

Dynamic Effects of Infant Industry Protection

Now suppose that the infant industry argument is valid and that by stimulating domestic production with a temporary import tariff, the domestic industry improves its own productive efficiency. We can represent this as a downward shift in the domestic industry supply curve. In actuality, this shift would probably occur gradually over time as the learning effects are incorporated in the production process. For analytical simplicity, we will assume that the effect occurs as follows. First, imagine that the domestic industry enjoys one period of protection in the form of a tariff. In the second period, we will assume that the tariff is removed entirely but that the industry experiences an instantaneous improvement in efficiency such that it can maintain production at its period one level but at the original free trade price. This efficiency improvement is shown as a supply curve shift from *S* to *S'* in [Figure 9.3 "Efficiency Improvement in a Small Importing Country"](#).

Figure 9.3 Efficiency Improvement in a Small Importing Country



This means that in the second period, free trade again prevails. The domestic price returns to the free trade price of P_1 , while domestic demand rises to D_1 . Because of the efficiency improvement, domestic supply in free trade is given by S_2 and the level of imports is $D_1 - S_2$ (the blue segment).

The static (one-period) welfare effects of the tariff removal and efficiency improvement are summarized in [Table 9.7 "Static Welfare Effects of Tariff Removal and Efficiency Improvement"](#). Note that these effects are calculated relative to the original equilibrium before the original tariff was implemented. We do this because we want to identify the welfare effects in each period relative to what would have occurred had the infant industry protection not been provided.

Table 9.7 Static Welfare Effects of Tariff Removal and Efficiency Improvement

	Importing Country
Consumer Surplus	0
Producer Surplus	+ E

	Importing Country
Govt. Revenue	0
National Welfare	+ E

Consumers again face the same free trade price that they would have faced if no protection had been offered. Thus they experience no loss or gain. Producers, however, face a new supply curve that generates a producer surplus of + E at the original free trade price. The government tariff is removed, so the government receives no tariff revenue. The net national welfare effect for the second period then is simply the gain in producer surplus.

The overall welfare impact over the two periods relative to no infant industry protection over two periods is simply the sum of each period's welfare effects. This corresponds to the sum of areas ($+ E - B - D$), which could be positive or negative. If the second-period producer surplus gain exceeds the first-period deadweight losses, then the protection has a positive two-period effect on national welfare.

But wait. Presumably the efficiency improvement in the domestic industry would remain, if not improve, in all subsequent periods as well. Thus it is not complete to consider the effects only over two periods. Instead, and for simplicity again, suppose that the new supply curve prevails in all subsequent periods. In this case, the true dynamic national welfare effects would consist of area E multiplied by the number of future periods we wish to consider minus the one-period deadweight losses. Thus even if the costs of the tariff are not made up in the second period, they may well be made up eventually at some point in the future. This would make it even more likely that the temporary protection would be beneficial in the long run.

If, in addition to the direct efficiency effects within the industry, there are spillover efficiency effects on other industries within the domestic economy, then the likelihood that temporary protection is beneficial is enhanced even further. In other words, over time, workers and managers from the protected industries may establish firms or take jobs in other sectors of the economy. Since they will bring their newly learned skills with them, it will cause an improvement in productive efficiency in those sectors as well. In this way, the supply of many manufacturing industries will be increased, allowing these sectors to compete more easily with firms in the rest of the world. Industrialization and GDP growth then is stimulated by the initial protection of domestic industries.

In summary, we have shown the possibility that protection of an infant industry may be beneficial for an economy. At the heart of the argument is the assumption that production experience generates efficiency improvements either directly in

the protected industry or indirectly in other industries as a learning spillover ensues. The infant industry argument relies on a dynamic view of the world rather than the static description used in classical trade models. Although protection may be detrimental to national welfare in the short run, it is conceivable that the positive dynamic long-run effects will more than outweigh the short-run (or static) effects.

The Economic Argument against Infant Industry Protection

The main economic argument against infant industry protection is that protection is likely to be a second-best policy choice rather than a first-best policy choice. The key element of the infant industry argument is the presence of a positive dynamic production externality. It is assumed that production experience causes learning, which improves future productive efficiency. Alternatively, it is assumed that these learning effects spill over into other industries and improve those industries' future productive efficiencies as well.

The theory of the second best states that in the presence of a market distortion, such as a production externality, it is possible to conceive of a trade policy that can improve national welfare. However, in this case, the trade policy—namely, the import tariff—is not the first-best policy because it does not attack the distortion most directly. In this case, the more-efficient policy is a production subsidy targeted at the industries that generate the positive learning effects.

To demonstrate this result, consider the following analytical example. We will use the same supply and demand conditions as depicted in [Figure 9.3 "Efficiency Improvement in a Small Importing Country"](#). The domestic supply and demand curves are given by D and S , respectively. The initial free trade world price of the good is P_1 . At that price, consumers would demand D_1 , but the domestic supply curve is too high to warrant any production. Thus the level of imports is given by D_1 .

Now suppose that the government implements a specific production subsidy equal to the difference in prices, $P_2 - P_1$. The subsidy would raise the producer price by the amount of the subsidy to P_2 , and hence domestic supply will rise to S_2 . The domestic consumer price would remain at P_1 , so demand would remain at D_1 . Imports would fall to $D_1 - S_2$.

The static (i.e., one-period) welfare effects of the production subsidy are shown in [Table 9.8 "Static Welfare Effects of a Production Subsidy"](#).

Table 9.8 Static Welfare Effects of a Production Subsidy

	Importing Country
Consumer Surplus	0
Producer Surplus	+ A
Govt. Revenue	- (A + B)
National Welfare	- B

Consumers of textiles are left unaffected by the subsidy since the domestic price remains the same. Producers gain in terms of producer surplus since the subsidy is sufficient to cause production to begin. In addition, employment is created in an industry. The government, however, must pay the subsidy. Thus someone pays higher taxes to fund the subsidy.

The net national welfare effect of the production subsidy is negative. Although some segments of the population benefit, there remains a production efficiency loss.

Note, however, that relative to an import tariff that generates the same level of domestic production, the subsidy is less costly in the aggregate. The production subsidy causes only a production efficiency loss, while the tariff causes an additional consumption efficiency loss. If the positive dynamic gains in efficiency in subsequent periods are the same, then the production subsidy would generate the same positive stream of benefits but at a lower overall cost to the country. For this reason, the production subsidy is the first-best policy to choose in light of the dynamic production externality. The import tariff remains second best.

For this reason, economists sometimes argue that although an import tariff may indeed be beneficial in the case of infant industries, it does not necessarily mean that protection is appropriate.

Other Arguments against Infant Industry Protection

Political economy problems. Political pressures in democratic economies can make it difficult to implement infant industry protection in its most effective manner. In order for protection to work in the long run, it is important that protection be temporary. There are two main reasons for this. First, it may be that the one-period efficiency improvement is less than the sum of the deadweight costs of protection. Thus if protection is maintained, then the sum of the costs may exceed the

efficiency improvements and serve to reduce national welfare in the long run. Second, and more critically, if protection were expected to be long lasting, then the protected domestic firms would have less incentive to improve their productive efficiency. If political pressures are brought to bear whenever the tariffs are scheduled to be reduced or removed, industry representatives might convince legislators that more time is needed to guarantee the intended efficiency improvements. In other words, firms might begin to claim that they need more time to compete against firms in the rest of the world. As long as legislators provide more time to catch up to world efficiency standards, protected firms have little incentive to incur the investment and training costs necessary to compete in a free market. After all, the tariff keeps the price high and allows even relatively inefficient production to produce profits for the domestic firms.

Thus one big problem with applying the infant industry protection is that the protection itself may eliminate the need for the firms to grow up. Without the subsequent efficiency improvements, protection would only generate costs for the economy in the aggregate.

Informational problems. In order for infant industry protection to work, it is important for governments to have reliable information about industries in their economies. They need to know which industries have strong learning effects associated with production and which industries are most likely to generate learning spillover effects to other industries. It would also be useful to know the size of the effects as well as the timing. But governments must decide not only which industries to protect but also how large the protective tariffs should be and over what period of time the tariff should be reduced and eliminated. If the government sets the tariff too low, the protection may be insufficient to generate very much domestic production. If the tariff is set too high, the costs of the tariff might outweigh the long-term efficiency improvements. If the tariff is imposed for too long a period, then firms might not have enough of an incentive to make the changes necessary to improve efficiency. If set for too short a time, then firms may not learn enough to compete with the rest of the world once the tariffs are removed.

Thus in order for infant industry protection to work, it is important to set the tariff for the correct industries, at the correct level, and for the correct period of time. Determining the correct industries, tariff level, and time period is not a simple matter. Indeed, some people argue that it is impossible to answer these questions with a sufficient amount of accuracy to warrant applying these policies.

Failure of import-substitution strategies. One popular development strategy in the 1950s and 1960s was known as import substitution. Essentially, this strategy is just

an application of the infant industry argument. However, many of the countries that pursued these kinds of inward-looking strategies, most notably countries in Latin America and Africa, performed considerably less well economically than many countries in Asia. The Asian countries—such as South Korea, Taiwan, Hong Kong, and Japan—pursued what have been labeled export-oriented strategies instead. Since many of these Southeast Asian countries performed so much better economically, it has lent some empirical evidence against the application of infant industry protection.

KEY TAKEAWAYS

- An import tariff that stimulates infant industry production sufficiently can raise national welfare over time, even for a small importing country.
- An import tariff is a second-best policy to correct for an infant industry production externality imperfection.
- A production subsidy is superior to an import tariff as a policy to correct for an infant industry production externality imperfection.
- In the presence of an infant industry production externality imperfection, a domestic policy is first best, while the best trade policy is second best.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe firms in less-developed countries that have a significant cost disadvantage compared with established firms located in the developed countries.
 - b. The type of comparative advantage that is not present in the short run but that develops in the long run.
 - c. The first-best policy option for a government that wishes to support an infant industry.
 - d. A second-best policy option for a government that wishes to support an infant industry.
 - e. Of *increase, decrease, no change, or ambiguous*, the effect of infant industry protection on national welfare under standard assumptions in the early periods while protection is in place.
 - f. Of *increase, decrease, no change, or ambiguous*, the effect of infant industry protection on national welfare under standard assumptions in the later periods after protection is removed.
 - g. Of *increase, decrease, no change, or ambiguous*, the effect of infant industry protection on overall national welfare under standard assumptions over all periods.

9.6 The Case of a Foreign Monopoly

LEARNING OBJECTIVES

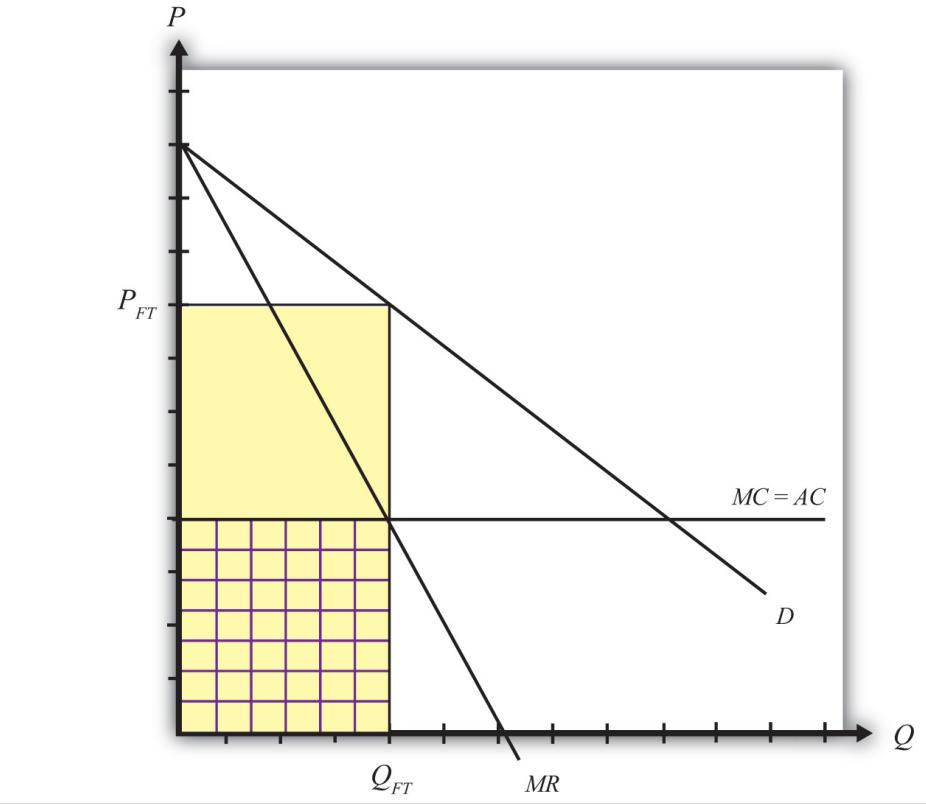
1. Learn that a foreign monopoly supplying products to domestic consumers is a type of market imperfection.
2. Recognize that a trade policy can be used to correct for a foreign monopolist imperfection.
3. Learn the first-best and second-best policy options to correct for a foreign monopolist imperfection.

Consider a domestic market supplied by a foreign monopoly firm. The domestic market consists of many consumers who demand the product but has no domestic producers of the product. All supply of the product comes from a single foreign firm.

Although this situation is not very realistic, it is instructive as an application of the theory of the second best. In this case, the market imperfection is that there are not a multitude of firms supplying the market. Rather, we have assumed the extreme opposite case of a monopoly supplier. To make this an international trade story, we simply assume the monopoly happens to be a foreign firm.

Consider the market described in [Figure 9.4 "Imports from a Foreign Monopoly Firm"](#). Domestic consumer demand is represented by a linear demand curve, D . When demand is linear, it follows that the marginal revenue curve will have twice the slope and will equal demand when the quantity is zero. Let the flat MC line represent a constant marginal cost in production for the foreign monopolist.

Figure 9.4 Imports from a Foreign Monopoly Firm



Assuming the monopolist maximizes profit, the profit-maximizing output level is found by setting marginal cost equal to marginal revenue. Why? Profit-maximizing output occurs at the quantity level Q_{FT} . At that quantity, the monopolist would set the price at P_{FT} , the only price that equalizes demand with its supply.

The monopolist's profit is the difference between total revenue and total cost. Total revenue is given by the product $(P_{FT}Q_{FT})$, the yellow area in the graph. Total cost is equal to average cost (AC) multiplied by output (Q_{FT}), given by the checkered area. The monopolist's profit is represented by the uncheckered yellow rectangular area in [Figure 9.4 "Imports from a Foreign Monopoly Firm"](#).

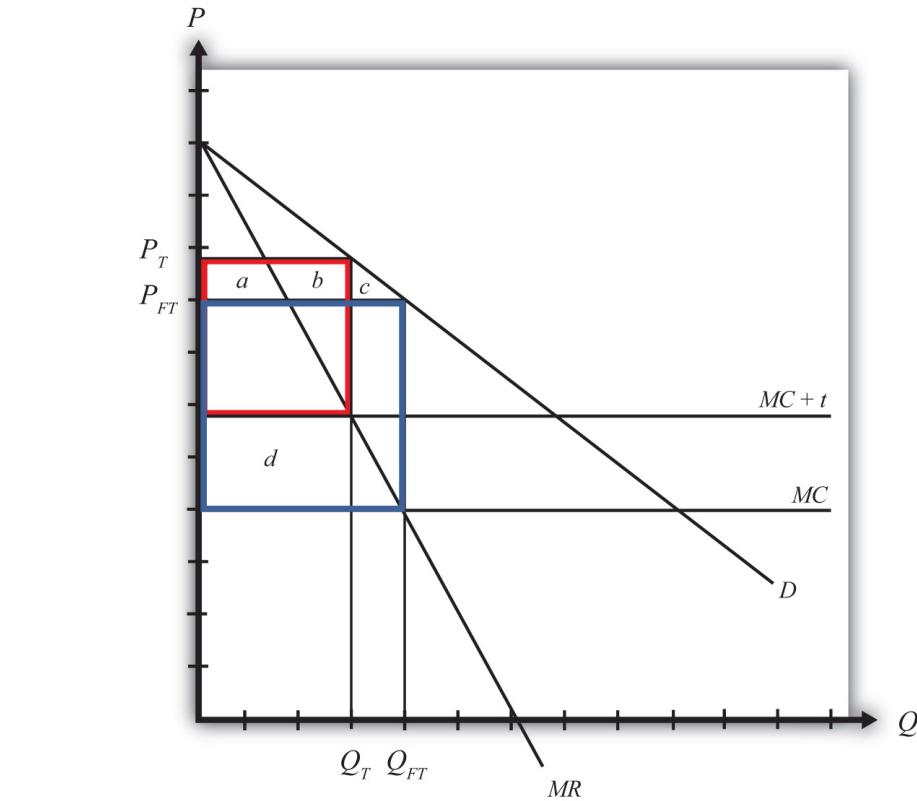
Strategic Trade Policy

Generally, strategic trade policy refers to cases of advantageous protection when there are imperfectly competitive markets. The case of a foreign monopolist represents one such case.

More specifically, though, the presence of imperfect competition implies that firms can make positive economic profit. Strategic trade policies typically involve the shifting of profits from foreign firms to domestic firms. In this way, national welfare can be improved, although it is often at the expense of foreign countries.

In this example, we shall consider the welfare effects of a specific tariff set equal to t . The tariff will raise the cost of supplying the product to the domestic market by exactly the amount of the tariff. We can represent this in [Figure 9.5 "A Tariff on Imports from a Foreign Monopoly Firm"](#) by shifting the marginal cost curve upward by the amount of the tariff to $MC + t$. The monopolist will reduce its profit-maximizing output to Q_T and raise its price to P_T . Note that the price rises by less than the amount of the tariff.

Figure 9.5 A Tariff on Imports from a Foreign Monopoly Firm



[Table 9.9 "Welfare Effects of a Tariff"](#) provides a summary of the direction and magnitude of the welfare effects to producers, consumers, and the government in the importing country as a result of the import tariff. The aggregate national welfare effects are also shown.

Table 9.9 Welfare Effects of a Tariff

	Importing Country
Consumer Surplus	- $(a + b + c)$
Producer Surplus	0
Govt. Revenue	+ d
National Welfare	$d - (a + b + c)$

Import tariff effects on the importing country's consumers. Consumers of the product in the importing country suffer a reduction in surplus because of the higher price that prevails. Refer to [Table 9.9 "Welfare Effects of a Tariff"](#) and [Figure 9.5 "A Tariff on Imports from a Foreign Monopoly Firm"](#) to see how the magnitude of the change in producer surplus is represented.

Import tariff effects on the importing country's producers. It is assumed that there are no domestic producers of the goods; thus there are no producer effects from the tariff.

Import tariff effects on the importing country's government. The government receives tariff revenue given by the per-unit tax (t) multiplied by the quantity of imports (Q_T). Who gains from the tariff revenue depends on how the government spends the money. Presumably these revenues help support the provision of public goods or help sustain transfer payments. In either case, someone in the economy ultimately benefits from the revenue. Refer to [Table 9.9 "Welfare Effects of a Tariff"](#) and [Figure 9.5 "A Tariff on Imports from a Foreign Monopoly Firm"](#) to see how the magnitude of the subsidy payments is represented.

The aggregate welfare effect for the importing country is found by summing the gains and losses to consumers, producers, and the government. The net effect consists of two components: a positive effect on the recipients of the government tariff revenue (d) and a negative effect on consumers ($a + b + c$), who lose welfare due to higher prices.

If demand is linear, it is straightforward to show that the gains to the country will always exceed the losses for some positive nonprohibitive tariff. In other words, there will exist a positive optimal tariff. Thus a tariff can raise national welfare when the market is supplied by a foreign monopolist.

One reason for this positive effect is that the tariff essentially shifts profits away from the foreign monopolist to the domestic government. Note that the original

profit level is given by the large blue rectangle shown in [Figure 9.5 "A Tariff on Imports from a Foreign Monopoly Firm"](#). When the tariff is implemented, the monopolist's profit falls to a level given by the red rectangle. Thus, in this case, the tariff raises aggregate domestic welfare as it reduces the foreign firm's profit.

First-Best Policy

Although a tariff can raise national welfare in this case, it is not the first-best policy to correct the market imperfection. A first-best policy must attack the imperfection more directly. In this case, the imperfection is the monopolistic supply of the product to the market. A monopoly maximizes profit by choosing an output level such that marginal revenue is equal to marginal cost. This rule deviates from what a perfectly competitive firm would do—that is, set price equal to marginal cost. When a firm is one among many, it must take the price as given. It cannot influence the price by changing its output level. In this case, the price is its marginal revenue. However, for a monopolist, which can influence the market price, price exceeds marginal revenue. Thus when the monopolist maximizes profit, it sets a price greater than marginal cost. This deviation—that is, $P > MC$ —is at the core of the market imperfection.

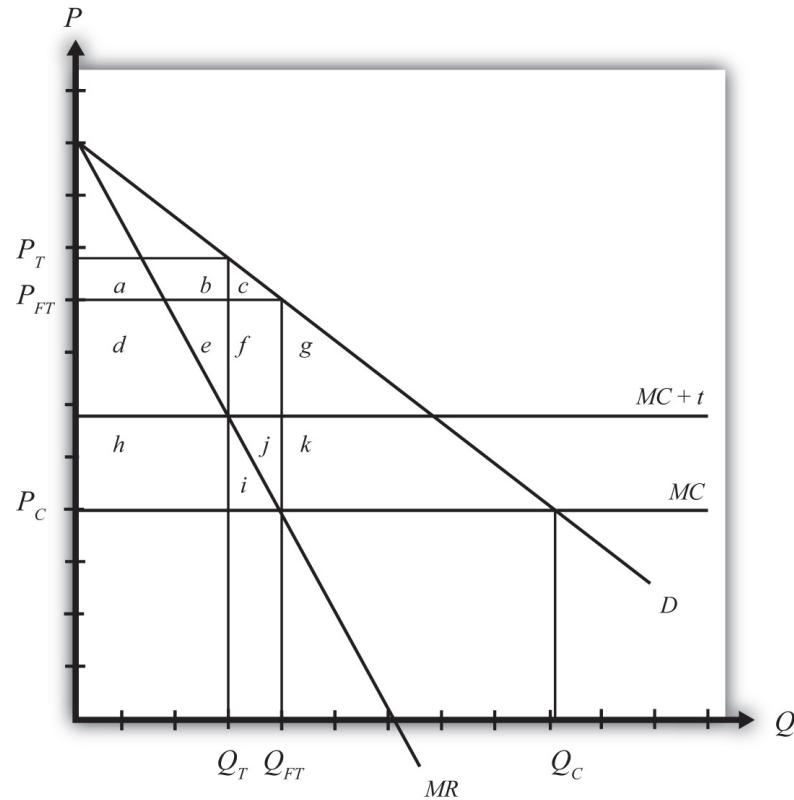
The standard way of correcting this type of imperfection in a domestic context is to regulate the industry. For example, electric utilities are regulated monopolies in the United States. Power can generally be purchased from only one company in any geographical area. To assure that these firms do not set exorbitant prices, the government issues a set of pricing rules that the firms must follow. The purpose is to force the firms to set prices closer, if not equal to, the marginal cost of production.

Now, in the case of utilities, determining the marginal cost of production is a rather difficult exercise, so the pricing rules to optimally regulate the industry are relatively complicated. In the case of a foreign monopolist with a constant marginal cost supplying a domestic market, however, the optimal policy is simple. The domestic government could merely set a price ceiling equal to the firm's marginal cost in production.

To see why a price ceiling is superior to a tariff, consider [Figure 9.6 "A Price Ceiling on Imports from a Foreign Monopoly Firm"](#). A second-best policy is the tariff. It would raise national welfare by the area $(h - a - b - c)$, which as mentioned will be positive for some tariffs and for a linear demand curve. The first-best policy is a price ceiling set equal to the marginal cost at P_C . The price ceiling would force the monopolist to set the price equal to the marginal cost and induce an increase in supply to Q_C . Consumers would experience an increase in consumer surplus, given

by the area $(d + e + f + g + h + i + j + k)$, because of the decline in price. Clearly, in this example, the consumer surplus gain with the price ceiling exceeds the national welfare gain from a tariff.

Figure 9.6 A Price Ceiling on Imports from a Foreign Monopoly Firm



This shows that although a tariff can improve national welfare, it is not the best policy to correct this market imperfection. Instead, a purely domestic policy—a price ceiling in this case—is superior.

KEY TAKEAWAYS

- A strategic trade policy attempts to shift foreign profits toward the importing economy.
- An import tariff levied against a foreign monopoly firm supplying domestic demand can raise national welfare.
- An import tariff is a second-best policy to correct for the imperfection of a foreign monopoly firm supplying domestic demand.
- A price ceiling is superior to an import tariff as a policy to correct for the imperfection of a foreign monopoly firm supplying domestic demand.
- In the presence of the imperfection of a foreign monopoly firm supplying domestic demand, a domestic policy is first best, while the best trade policy is second best.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The first-best policy option for a government that faces a foreign monopoly (with constant marginal costs) as the sole firm selling a product in the domestic market.
 - b. A second-best policy for a government that faces a foreign monopoly (with constant marginal costs) as the sole firm selling a product in the domestic market.
 - c. The term used to describe a policy that shifts profits from foreign firms toward groups in the domestic economy.
2. Suppose the U.S. market demand for VCRs is given by $D = 1,000 - 2P$. The U.S. market is supplied by a foreign monopolist with a constant marginal cost of production equal to \$200. The marginal revenue curve faced by the supplier is given by $MR = 500 - Q$.
 - a. Calculate the equilibrium price and quantity of imports of VCRs. Depict this equilibrium graphically.
 - b. Calculate consumer surplus in this market equilibrium.

Suppose the government imposes a specific tariff of \$100.

 - c. Calculate the new equilibrium price and quantity.
 - d. Calculate the change in consumer surplus and the tariff revenue.
 - e. What is the change in national welfare?
 - f. What is the first-best policy action to raise national welfare in this case? If this policy is applied, what would be the domestic price and quantity imported?
 - g. Calculate the change in national welfare if the first-best policy is applied rather than the tariff. Compare this with the national welfare effect of the tariff.
 - h. Briefly explain how to identify first-best policies in general and explain why the policy in this case satisfies the criterion.

9.7 Monopoly and Monopsony Power and Trade

LEARNING OBJECTIVES

1. Learn that monopoly power and monopsony power in trade are types of market imperfections.
2. Recognize that a trade policy can be used to correct for a large-country imperfection.
3. Learn the first-best and second-best policy options to correct for a large-country imperfection.

Perhaps surprisingly, “large” importing countries and “large” exporting countries have a market imperfection present. This imperfection is more easily understood if we use the synonymous terms for “largeness”: monopsony power and monopoly power. Large importing countries are said to have “monopsony power in trade,” while large exporting countries are said to have “monopoly power in trade.” As this terminology suggests, the problem here is that the international market is not perfectly competitive. For complete perfect competition to prevail internationally, we would have to assume that all countries are “small” countries.

Let’s first consider monopoly power. When a large exporting country implements a trade policy, it will affect the world market price for the good. That is the fundamental implication of largeness. For example, if a country imposes an export tax, the world market price will rise because the exporter will supply less. It was shown in [Chapter 7 "Trade Policy Effects with Perfectly Competitive Markets"](#), [Section 7.23 "Export Taxes: Large Country Welfare Effects"](#) that an export tax set optimally will cause an increase in national welfare due to the presence of a positive terms of trade effect. This effect is analogous to that of a monopolist operating in its own market. A monopolist can raise its profit (i.e., its firm’s welfare) by restricting supply to the market and raising the price it charges its consumers. In much the same way, a large exporting country can restrict its supply to international markets with an export tax, force the international price up, and create benefits for itself with the terms of trade gain. The term monopoly “power” is used because the country is not a pure monopoly in international markets. There may be other countries exporting the product as well. Nonetheless, because its exports are a sufficiently large share of the world market, the country can use its trade policy in a way that mimics the effects caused by a pure monopoly, albeit to a lesser degree. Hence the country is not a monopolist in the world market but has monopoly “power” instead.

Similarly, when a country is a large importer of a good, we say that it has “monopsony power.” A monopsony is a single buyer in a market consisting of many sellers. A monopsony raises its own welfare or utility by restricting its demand for the product and thereby forcing the sellers to lower their price. By buying fewer units at a lower price, the monopsony becomes better off. In much the same way, when a large importing country places a tariff on imports, the country’s demand for that product on world markets falls, which in turn lowers the world market price. It was shown in [Chapter 7 "Trade Policy Effects with Perfectly Competitive Markets"](#), [Section 7.6 "The Optimal Tariff"](#) that an import tariff, set optimally, will raise national welfare due to the positive terms of trade effect. The effects in these two situations are analogous. We say that the country has monopsony “power” because the country may not be the only importer of the product in international markets, yet because of its large size, it has the “power” of a pure monopsony.

First-Best or Second-Best Trade Policies

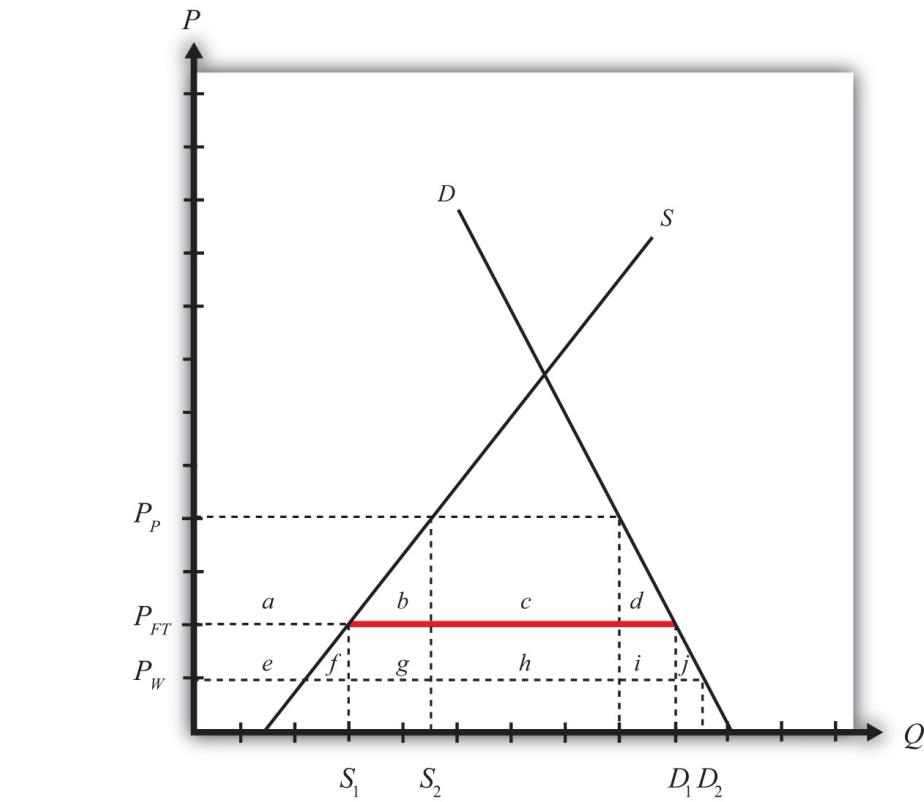
It has already been shown that a trade policy can improve a country’s national welfare when that country is either a large importer or a large exporter. The next question to ask is whether the optimal tariff or the optimal export tax, each of which is the very best “trade” policy that can be chosen, will raise national welfare to the greatest extent or whether there is another purely domestic policy that can raise welfare to a larger degree.

Because a formal graphical comparison between the first-best and second-best policies is difficult to construct in this case, we will rely on an intuitive answer based on what has been learned so far. It is argued in [Chapter 9 "Trade Policies with Market Imperfections and Distortions"](#), [Section 9.3 "The Theory of the Second Best"](#) that the first-best policy will always be that policy that attacks the market imperfection or market distortion most directly. In the case of a large country, it is said that the market imperfection is a country’s monopsony or monopoly power. This power is exercised in “international” markets, however. Since benefits accrue to a country by changing the international terms of trade in a favorable direction, it is through trade that the monopsony or monopoly power can “best” be exercised. This observation clearly indicates that trade policies will be the first-best policy options. When a country is a large importing country, an optimal tariff or import quota will be first best. When a country is a large exporting country, an optimal export tax or voluntary export restraint (VER) will be first best.

Now, of course, this does not mean that a purely domestic policy cannot raise national welfare when a country is “large.” In fact, it was shown in [Chapter 8 "Domestic Policies and International Trade"](#), [Section 8.4 "Production Subsidy Effects in a Small Importing Country"](#) that an import tariff is equivalent to a domestic production subsidy and a domestic consumption tax set at the same level;

thus setting one of these policies at an appropriate level may also be able to raise national welfare. To see that this is true, let's consider a large importing country initially in free trade. Because it is in free trade, there is a market imperfection present that has not been taken advantage of. Suppose this country's government implements a production subsidy provided to the domestic import-competing firm. We can work out the effects of this production subsidy in [Figure 9.7 "Domestic Production Subsidy by a Large Importing Country"](#).

Figure 9.7 Domestic Production Subsidy by a Large Importing Country



The free trade price is given by P_{FT} . The domestic supply in free trade is S_1 , and domestic demand is D_1 , which determines imports in free trade as $D_1 - S_1$ (the red line in [Figure 9.7 "Domestic Production Subsidy by a Large Importing Country"](#)).

When a specific production subsidy is imposed, the producer's price rises, at first by the value of the subsidy. The consumer's price is initially unaffected. This increase in the producer's price induces the producer to increase its supply to the market. The supply rises along the supply curve and imports begin to fall. However, because the country is a large importer, the decrease in imports represents a decrease in the world demand for the product. As a result, the world price of the good falls, which

in turn means that the price paid by consumers in the import market also falls. When a new equilibrium is reached, the producer's price will have risen (to P_P in [Figure 9.7 "Domestic Production Subsidy by a Large Importing Country"](#)), the consumer's price will have fallen (to P_W), and the difference between the producer and consumer prices will be equal to the value of the specific subsidy ($s = P_P - P_W$). Note that the production subsidy causes an increase in supply from S_1 to S_2 and an increase in demand from D_1 to D_2 . Because both supply and demand rise, the effect of the subsidy on imports is, in general, ambiguous.

The welfare effects of the production subsidy are shown in [Table 9.10 "Welfare Effects of a Production Subsidy in a Large Country"](#). The letters refer to the area in [Figure 9.7 "Domestic Production Subsidy by a Large Importing Country"](#).

Table 9.10 Welfare Effects of a Production Subsidy in a Large Country

Consumer Surplus	$+ (e + f + g + h + i + j)$
Producer Surplus	$+ \alpha$
Govt. Revenue	$- (\alpha + b + e + f + g)$
National Welfare	$h + i + j - b$

The first thing to note is that the production subsidy causes welfare improvements for both producers and consumers. All previous policies have these two groups always experiencing opposite effects. It would appear, in this case, we have struck the “mother lode”—finally, a policy that benefits both consumers and producers. Of course, the effects are not all good. To achieve this effect, the government must pay the subsidy to the firms, and that must come from an increase in taxes either now or in the future. So the country must incur a cost in the form of government expenditures. The final effect—that is, the effect on national welfare—is ambiguous. However, it is conceivable that the area given by $(h + i + j)$ may exceed the area (b) , in which case, national welfare will rise. Of course, if a different subsidy level is set, it is also possible that national welfare will fall. It will depend on the value of the subsidy, and it will vary across every separate market.

In the case that welfare does rise, it will occur because the country is a large importer. The domestic production subsidy allows the country to take advantage of its monopsony power in trade. By stimulating domestic production, the subsidy reduces import demand, which pushes the price of the country's import good down in the world market. In other words, the country's terms of trade improves. In this way, a country can take advantage of its monopsony power by implementing a

domestic policy, such as a production subsidy to an import-competing industry. Note well, though, that not every subsidy provided will raise national welfare. The subsidy must be set at an appropriate level for the market conditions to assure an increase in national welfare. In general, a relatively small subsidy will achieve this objective. If the subsidy is set too high, the losses from government expenditures will exceed the gains to consumers and producers, and the country will suffer national welfare losses.

Other domestic policies can also be used to raise national welfare in the case of a large importing country. Indeed, any policy that restricts international demand for a product will potentially raise national welfare—only “potentially” because it is necessary to set the policy at the proper level. The other obvious domestic policy that can achieve this result is a domestic consumption tax on the imported product. Recall that a consumption tax is one of the two domestic policies that, when applied together, substitutes for an import tariff. Since the import tariff can raise welfare, so can its constituent parts.

KEY TAKEAWAYS

- A market imperfection exists whenever a country is “large”: either a large importer, a large exporter, or both.
- In these cases, international perfect competition does not prevail. We say that a large exporting country has monopoly power in trade, while a large importing country has monopsony power in trade.
- Due to the presence of the market imperfection, a trade policy can raise the nation’s welfare above the level possible with free trade.
- Domestic policies, such as production subsidies and consumption taxes, can also raise national welfare when a country is large.
- The first-best policy in the case of a large country is a trade policy.
- A trade policy most directly attacks the market distortion—that is, international imperfect competition.
- If a country is a large importer, the first-best trade policy is the optimal tariff or its equivalent quota.
- If a country is a large exporter, the first-best policy is the optimal export tax or its equivalent VER.
- Domestic policies, used alone, are second-best policy options.

EXERCISES

- Consider the following imperfect market situation in the table below. From the following list of policy options, identify all types of trade policies *and* all types of domestic policies that could potentially raise national welfare in the presence of each imperfection. Consider only the partial equilibrium effects of each policy.

Options: An import tariff, an import quota, a voluntary export restraint (VER), an export tax, an export subsidy, a production tax, a production subsidy, a consumption tax, and a consumption subsidy.

TABLE 9.11 WELFARE IMPROVING POLICIES

	Trade Policy	Domestic Policy
A large country that imports steel		

- Consider the domestic policy action listed along the top row of the table below. In the empty boxes, use the following notation to indicate the effect of the policy on the variables listed in the first column. Use a partial equilibrium model to determine the answers and assume that the shapes of the supply and demand curves are “normal.” Assume that the policy does not begin with, or result in, prohibitive policies. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

TABLE 9.12 EFFECTS OF A PRODUCTION SUBSIDY

	Production Subsidy by a Large Importing Country
Domestic Consumer Price	
Domestic Producer Price	
Domestic Consumer Welfare	
Domestic Producer Welfare	
Domestic Government Revenue	
Domestic National Welfare	
Foreign Price	
Foreign National Welfare	

9.8 Public Goods and National Security

LEARNING OBJECTIVES

1. Learn that public goods, which have the features of being nonrival and nonexcludable in consumption, are a type of market imperfection.
2. Recognize that a trade policy can be used to correct for a public good imperfection.
3. Learn the first-best and second-best policy options to correct for a public good imperfection.

One of the oldest and most common arguments supporting protection is the “national security argument,” also called the “national defense argument.” This argument suggests that it is necessary to protect certain industries with a tariff to assure continued domestic production in the event of a war. Many products have been identified as being sufficiently important to warrant protection for this reason. Perhaps the most common industry identified is agriculture. Simply consider the problems that would arise if a nation did not have an adequate food supply when it was at war with the outside world. Low food stocks may induce severe hardships and even famine. A simple solution to avoid this potential problem is to maintain a sufficiently high tariff in order to keep cheap foreign goods out and, in turn, maintain production of the domestic goods.

Similar problems may arise in many other industries. Consider the potential problems for a country’s national security if it could not produce an adequate amount of steel, aluminum, ships, tanks, planes, fuel, and so on in the event of a war. The number of products that could be added to this list is enormous. Indeed, at one time or another in most countries’ histories, it has been argued that almost every product imaginable is important from a national security perspective and thus is deserving of protection. One of the most interesting arguments ever described is that made by the embroidery industry, which once argued for a protective tariff in the United States because embroidered patches on soldiers’ uniforms are essential in maintaining the morale of the troops. Thus it was clear, to them at least, that the embroidery industry needed to be protected for national security reasons.

National Security and Public Goods

We can make better sense of the national security argument if we classify it in the context of the theory of the second best. In this case, we must note that the national security argument is actually incorporating a market imperfection into the story to justify the use of a protective tariff. The market imperfection here is a public good. National security is a public good and public goods are excluded from the standard assumptions of perfect competition. Thus, whenever a product has public good characteristics, we can say that a market imperfection is present. Traditionally, the literature in economics refers to concerns such as national security as noneconomic objectives. The effects that food production may have on the nation's sense of security, for example, were thought to fall outside the realm of traditional economic markets.

In general, public goods have the following two consumption characteristics: they are nonexcludable and they are nonrival. Nonexcludability means that once the product is produced, it is impossible to prevent people from consuming it. Nonrivalry means that many people can consume the produced product without diminishing its usefulness to others. Here are a few examples to explain the point. First, consider a nonpublic good: soda. A soda is excludable since the producer can put it into a can and require you to pay for it to enjoy its contents. A can of soda is also a rival good. That's because if you consume the can of soda, there is no way for anyone else to consume the *same* can. This implies that a can of soda is *not* a public good. On the other hand, consider oxygen in the atmosphere. (This is an odd example because oxygen in the air is not formally produced, but let's ignore that for a moment.) Atmospheric oxygen is nonexcludable because once it is there, everyone has free access to its use. It is impossible (or at least very difficult) to prevent some people from enjoying the benefits of the air. Atmospheric oxygen is also nonrival because when one person takes a breath, it does not diminish the usefulness of the atmosphere for others. Thus, if atmospheric oxygen did need to be formally produced, it would be a classic example of a pure public good.

The typical examples of public goods include national security, clean air, lighthouse services, and commercial-free television and radio broadcasts. National security is the public good we are most concerned with in international trade. It is a public good because, once provided, (1) it is difficult to exclude people within the country from the safety and security generated and (2) multiple individuals can enjoy the added safety and security without limiting that received by others.

We know from the theory of the second best that when market imperfections are present, government policies can be used to improve the national welfare. In most cases, trade policies can be used as well. It is well known in economic theory that when a good has public good characteristics, and if private firms are free to supply

this good in a free market, then the public good will *not* be adequately supplied. The main problem occurs because of free ridership. If a person believes that others may pay for a good and if its subsequent provision benefits all people—due to the two public good features—then that person may avoid paying for the good in a private marketplace. If many people don't pay, then the public good will be insufficiently provided relative to the true demands in the country. It is well known that government intervention can solve this problem. By collecting taxes from the public, and thus forcing everyone to pay some share of the cost, the public good can be provided at an adequate level. Thus national welfare can be increased with government provision of public goods.

A similar logic explains why a trade policy can be used to raise a country's welfare in the presence of a public good. It is worth pointing out, though, that the goods highlighted above, such as agricultural products and steel production, are not themselves public goods. The public good one wishes to provide in greater abundance is “national security.” And it is through the production of certain types of goods locally that more security can be provided. For example, suppose it is decided that adequate national security is possible only if the nation can provide at least 90 percent of its annual food supplies during wartime. Suppose also that under free trade and laissez-faire domestic policies, the country produces only 50 percent of its annual food supply and imports the remaining 50 percent. Finally, suppose the government believes that it would be very difficult to raise domestic production rapidly in the event that imported products were ever cut off, as might occur during a war. In this case, a government may decide that its imports are too high and thus pose a threat to the country's national security.

A natural response in this instance is to put high tariffs in place to prevent imports from crowding out domestic production. Surely, a tariff exists that will reduce imports to 10 percent and subsequently cause domestic production to rise to 90 percent. We know from tariff analysis that in the case of a small country, a tariff will cause a net welfare loss for the nation in a perfectly competitive market. These same gains and losses and net welfare effects can be expected to prevail here. However, because of the presence of the public good characteristics of national security, there is more to the story. Although the tariff alone causes a net welfare loss for the economy, the effect is offset with a positive benefit to the nation in the form of greater security. If the added security adds more to national welfare than the economic losses caused by the tariff, then overall national welfare will rise. Thus protectionism can be beneficial for the country.

The national security argument for protection is perfectly valid and sound. It is perfectly logical under these conditions that protectionism can improve the nation's welfare. However, because of the theory of the second best, many economists remain opposed to the use of protectionism, even in these

circumstances. The reason is that protectionism turns out to be a second-best policy option.

Recall that the first-best policy response to a market imperfection is a policy that is targeted as directly as possible at the imperfection itself. Thus, if the imperfection arises because of some production characteristic, a production subsidy or tax should be used. If the problem is in the labor market, a tax or subsidy in that market would be best, and if the market imperfection is associated with international trade, then a trade policy should be used.

In this case, one might argue that the problem is trade related, since one can say that national security is diminished because there are too many imports of, say, agricultural goods. Thus an import tariff should be used. However, this logic is wrong. The actual problem is maintaining an adequate food supply in a time of war. The problem is really a production problem because if imports were to be cut off in an emergency, the level of production would be too low. The most cost-effective way, in this situation, to maintain production at adequate levels will be a production subsidy. The production subsidy will raise domestic production of the good and can be set high enough to assure that an adequate quantity is produced each year. The subsidy will cost the government money and it will generate a net production efficiency loss. Nevertheless, the efficiency loss from a tariff, one that generates the same level of output as a production subsidy, will cause an even greater loss. This is because an import tariff generates both a production efficiency loss and a consumption efficiency loss. Thus, to achieve the same level of production of agricultural goods, a production subsidy will cost less overall than an import tariff. We say, then, that an import tariff is a second-best policy. The first-best policy option is a production subsidy.

Another Case in Which a Trade Policy Is First Best

There is one case in which a trade policy, used to protect or enhance national security, is the first-best policy option. Consider a country that produces goods that could be used by other countries to attack or harm the first country. An example would be nuclear materials. Some countries use nuclear power plants to produce electricity. Some of the products used in this production process, or the knowledge gained by operating a nuclear facility, could be used as an input in the production of more dangerous nuclear weapons. To prevent such materials from reaching countries, especially materials that may potentially threaten a country, export bans are often put into place. The argument to justify an export ban is that preventing certain countries from obtaining materials that may be used for offensive military purposes is necessary to maintain adequate national security.

In the United States, export bans are in place to prevent the proliferation of a variety of products. Many other products require a license from the government to export the product to certain countries. This allows the government to monitor what is being exported to whom and gives them the prerogative to deny a license if it is deemed to be a national security threat. In the United States, licenses are required for goods in short supply domestically; goods related to nuclear proliferation, missile technology, and chemical and biological weapons; and other goods that might affect regional stability, crime, or terrorist activities. In addition, the United States maintains a Special Designated Nationals list, which contains names of organizations to which sales of products are restricted, and a Denied Persons list, which contains names of individuals with whom business is prohibited. In recent years the United States has maintained export bans to several countries, including Cuba, Iran, Syria, and Sudan.

In this case, the export control policy is the first-best policy to enhance national security. This is because the fundamental problem is certain domestic goods getting into the hands of certain foreign nations, groups, or individuals. The problem is a *trade problem* best corrected with a *trade policy*. Indeed, there is no effective way to control these sales, and thus to enhance national security, using a purely domestic policy.

KEY TAKEAWAYS

- The preservation of national security is a common justification for the use of protection.
- The preservation of national security is a type of noneconomic objective.
- Protection can help maintain an adequate domestic supply of materials critical in the event of war, including food, steel, military equipment, and petroleum.
- Export bans can be used to prevent the proliferation of materials that may eventually prove to be threatening to a nation's security.
- Import tariffs can raise national welfare when increased production of the protected product enhances national security.
- Because national security is a public good and also an imperfection, trade protection can sometimes be beneficial for a country.
- A production subsidy can achieve the same level of production at a lower cost.
- A production subsidy is the first-best policy when increased production of a good enhances national security.
- An import tariff is a second-best policy option.
- An export ban can raise a nation's welfare when the export of a product reduces national security.
- The export ban, a trade policy, is the first-best policy option when export of a product reduces national security.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe policy intentions that are not economic in nature.
 - b. This is a common justification for import protection of food, steel, shipping, and many other things thought necessary under certain circumstances.
 - c. This policy is first best if a product in the hands of foreigners could threaten one’s national security.
 - d. Of *a production subsidy* or *an import tariff*, this policy is likely to be first best to protect a nation’s agricultural production.
 - e. The term describing a “good” like national security that is both nonexcludable and nonrival in consumption.

9.9 Trade and the Environment

LEARNING OBJECTIVES

1. Learn that environmental externalities are a type of market imperfection.
2. Recognize that a trade policy can be used to correct for an environmental imperfection.
3. Learn the first-best and second-best policy options to correct for an environmental imperfection.

One contentious issue in international trade policy discussions concerns the connection between international trade and the environment. Many environmental groups claim that freer trade, as implemented through the World Trade Organization (WTO) agreements or in free trade agreements such as the North American Free Trade Agreement (NAFTA), results in negative environmental outcomes. For example, the Sierra Club argues, “Economic globalization ties the world together as never before. But it also poses serious new threats to our health and the environment. Trade agreements promote international commerce by limiting governments’ ability to act in the public interest. Already food safety, wildlife and pollution control laws have been challenged and weakened under trade rules as illegal ‘barriers to trade.’” Sierra Club, “A Fair Trade Bill of Rights,” Responsible Trade, <http://www.sierraclub.org/trade/ftaa/rights.asp>.

In contrast, the WTO, a frequent target for criticism by environmental groups, points to the WTO agreement, which states, “[WTO member] relations in the field of trade and economic endeavor should be conducted with a view to raising standards of living...while allowing for the optimal use of the world’s resources in accordance with the objective of sustainable development, seeking both to protect and preserve the environment and to enhance the means for doing so in a manner consistent with their respective needs and concerns at different levels of economic development.” World Trade Organization, “Environment Issues: Sustainable development,” http://www.wto.org/english/tratop_e/envir_e/sust_dev_e.htm.

Arguably, the stated goals of free trade-oriented groups and environmental groups are very similar, at least as highlighted in the documents produced by both sides. What differ are the methods used to achieve the objectives. For reasons to be elucidated below, the WTO has argued that environmental concerns are not directly within the purview of the WTO agreement, but despite that, environmental policies

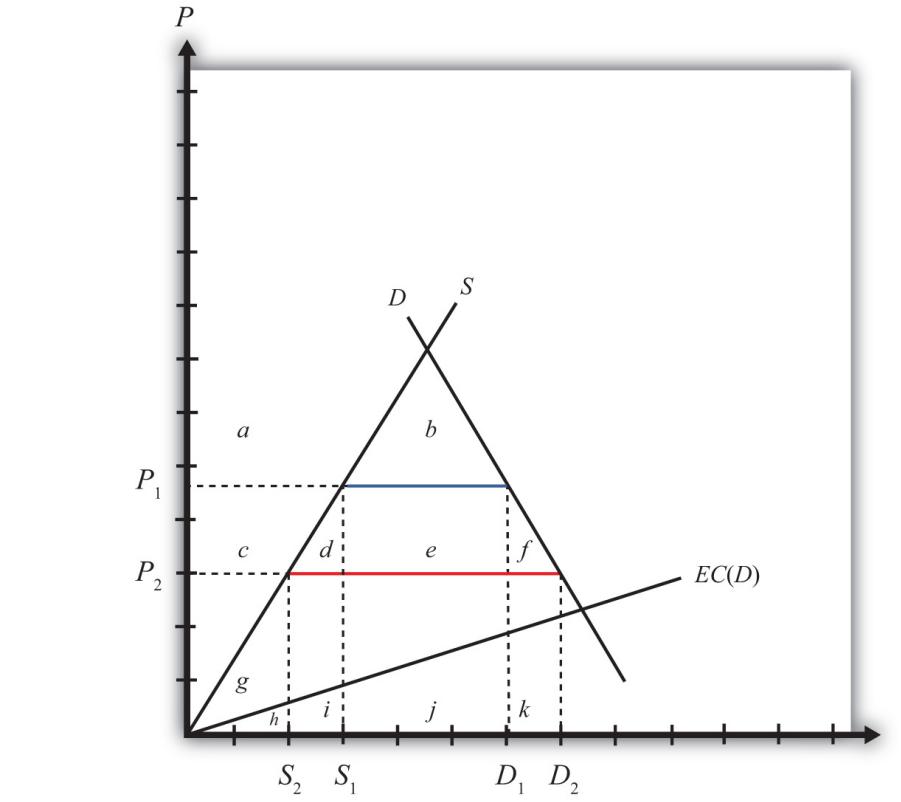
and international environmental agreements are neither prohibited by nor inconsistent with the WTO accords. In essence, the argument by some has been that the WTO agreement, and free trade agreements more generally, is intended to be about trade and is not intended to solve tangential problems related to the environment. On the other hand, environmental groups have pointed out that sometimes WTO and free trade agreement decisions have a negative effect on environmental outcomes, and thus these agreements should be revised to account for these negative effects.

Below we will consider these issues with respect to one type of environmental concern: pollution caused by consumption of an imported good. Although we will not consider many of the other contested environmental and trade issues, this one example will suffice to establish some important and generalizable conclusions.

Trade Liberalization with Environmental Pollution

Consider a small country importing gasoline with a tariff in place initially such that the domestic tariff-inclusive price is P_1 . At this price, domestic supply is S_1 , domestic demand (or consumption) is D_1 , and the level of imports is $(D_1 - S_1)$, shown in [Figure 9.8 "Trade and Environmental Costs"](#).

Figure 9.8 Trade and Environmental Costs



Suppose that domestic consumption of gasoline causes air pollution. This means that consumption has a negative external effect on all users of air—that is, there is a negative consumption externality.

Let's assume that the cost to society (in dollar terms) of the air pollution is an increasing function of domestic consumption. In other words, the greater the consumption of gasoline, the greater is the pollution, and the greater is the subsequent harm caused to people in the country. For simplicity, assume the environmental cost, $EC(D)$, is a linear function of total domestic demand, D . The height of EC at any level of demand represents the additional dollar cost of an additional gallon of gasoline consumption. This implies the total environmental cost of a consumption level—say, D_2 —is the area under the EC curve between the origin and D_2 .

With the initial tariff in place, domestic demand is D_1 , which implies that the total societal cost of pollution is given by the area $(h + i + j)$. Note that despite the cost of pollution, it does make sense to produce and consume this good if the objective is national welfare. Consumer surplus is given by the area $(a + b)$ and producer surplus

is $(c + g)$. The sum of these two clearly exceeds the social cost of pollution, $(h + i + j)$. (Note that these statements are true for [Figure 9.8 "Trade and Environmental Costs"](#) in particular; they are not true in general. By drawing the EC curve very steeply, corresponding to a much higher cost of pollution, it might not make sense to produce and consume the good in the market equilibrium.)

Next, suppose that the country agrees to remove the tariff on imported gasoline after signing a trade liberalization agreement. The question we ask is, Can trade liberalization have such a negative effect on the environment that it makes a country worse off? The answer, as we'll see, is yes.

Suppose the tariff is removed and the price of gasoline falls to P_2 . The lower price causes a reduction in production to S_2 , an increase in consumption to D_2 , and an increase in imports from the blue line segment ($D_1 - S_1$) to the red line segment ($D_2 - S_2$). Since domestic consumption of gasoline rises, there is also an increase in pollution.

The welfare effects of the tariff elimination are summarized in [Table 9.13 "Welfare Effects of a Tariff Elimination with a Negative Environmental Consumption Externality"](#). The letters refer to the area in [Figure 9.8 "Trade and Environmental Costs"](#).

Table 9.13 Welfare Effects of a Tariff Elimination with a Negative Environmental Consumption Externality

	Importing Country
Consumer Surplus	$+ (c + d + e + f)$
Producer Surplus	$- c$
Govt. Revenue	$- e$
Pollution Effect	$- k$
National Welfare	$(d + f) - k$

Consumers of gasoline benefit by the areas $(c + d + e + f)$ from the lower free trade price. Domestic producers lose (c) with a reduction in producer surplus. The government also loses tariff revenue (e) . The net total efficiency gains from trade are given by the areas $(d + f)$. However, the presence of the environmental consumption externality means there is an additional cost (k) caused by the pollution from higher domestic consumption of gasoline.

The national welfare effect of the tariff elimination is given by $(d + f - k)$. For a particular level of efficiency gains, the total national effect will depend on the size of the pollution cost. In the graph, the curves are drawn such that area k is slightly larger than $d + f$. Thus trade liberalization can cause a reduction in national welfare. The cost of additional pollution may be greater than the efficiency improvements from free trade. However, if the environmental cost of consumption were lower, the $EC(D)$ line would be flatter and area k would become smaller. Thus for lower environmental costs, trade liberalization might raise national welfare. The net effect, positive or negative, will depend on the magnitude of the pollution costs relative to the efficiency benefits.

Trade Policy versus Domestic Policy

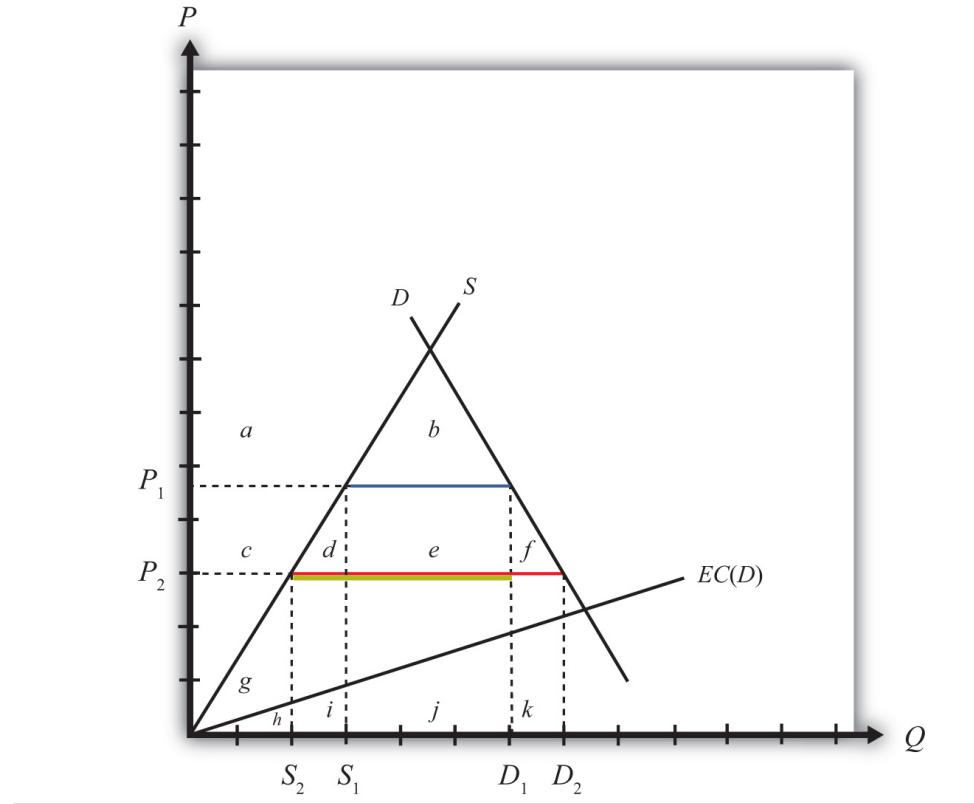
In general, the theory of the second best suggests that, in the presence of a market imperfection or distortion, a properly chosen trade policy might be found that will raise a small country's national welfare. However, for most imperfections, a trade policy will be a second-best policy. A better policy, a first-best policy, will always be that policy that attacks the imperfection or distortion most directly. In most instances, the first-best policy will be a domestic policy rather than a trade policy.

In this case, environmental pollution caused by the consumption of gasoline is a market imperfection because gasoline consumption has a negative external effect (via pollution) on others within the society. Economists call this a negative consumption externality. This problem can be corrected with any policy that reduces the negative effect at a cost that is less than the benefit. A tariff on imports is one such policy that could work. However, the most direct policy option, hence the first-best policy choice, is a consumption tax. Below we'll show the welfare effects of a tariff and a domestic consumption tax and compare the results to demonstrate why a consumption tax is first best while a tariff is second best.

Welfare Effects of a Tariff with Environmental Pollution

First, let's consider the effects of a tariff when consumption of the import good causes pollution. Consider a small country importing gasoline at the free trade price given by P_2 in [Figure 9.9 "Tariff Effects and Environmental Costs"](#). (Note that this is [Figure 9.8 "Trade and Environmental Costs"](#) redrawn.) Demand is given by D_2 , supply by S_2 , and imports are $(D_2 - S_2)$ (the red line). Suppose that domestic consumption of gasoline causes air pollution. Assume the environmental cost of pollution in dollar terms, $EC(D)$, is a linear function of total domestic demand, D .

Figure 9.9 Tariff Effects and Environmental Costs



Next, suppose a specific tariff, $t = P_1 - P_2$, is imposed, thereby raising the domestic price to P_1 . Domestic demand for gasoline falls to D_1 , supply rises to S_1 , and imports fall to $(D_1 - S_1)$ (the blue line). The welfare effects of the tariff are presented in [Table 9.14 "Welfare Effects of a Tariff with a Negative Environmental Consumption Externality"](#). The letters refer to the areas in [Figure 9.9 "Tariff Effects and Environmental Costs"](#).

Table 9.14 Welfare Effects of a Tariff with a Negative Environmental Consumption Externality

	Importing Country
Consumer Surplus	$-(c + d + e + f)$
Producer Surplus	$+ c$
Govt. Revenue	$+ e$
Pollution Effect	$+ k$
National Welfare	$k - (d + f)$

Consumers of gasoline lose surplus ($c + d + e + f$) from the higher domestic price. Domestic producers gain (+c) with an increase in producer surplus. The government also collects tariff revenue (+e). The net total efficiency losses from trade are given by the areas ($d + f$). However, the presence of the environmental consumption externality means there is an additional benefit caused by the reduced pollution. This benefit is represented by the area k .

The net national welfare effect of the tariff is given by ($k - d - f$). Since the curves are drawn such that area k is slightly larger than $d + f$, a tariff results in an improvement in national welfare in this example. More generally, we can only say that a tariff may result in an increase in national welfare since it will depend on the shapes of the curves and the size of the tariff.

Welfare Effects of a Consumption Tax with Environmental Pollution

Next, suppose that a consumption tax, $t = P_1 - P_2$, is imposed instead of a tariff. Refer to [Figure 9.9 "Tariff Effects and Environmental Costs"](#). The tax will raise the consumer's price to P_1 but will leave the producer's price at P_2 . Domestic producers will not be affected by the consumption tax since continued competition in free trade with firms in the rest of the world will maintain their profit-maximizing price at the world price of P_2 . The price changes will cause domestic demand for gasoline to fall to D_1 , but supply will remain at S_2 . Imports will fall to $(D_1 - S_2)$ (the yellow line). The welfare effects of the consumption tax are presented in [Table 9.15 "Welfare Effects of a Domestic Consumption Tax with a Negative Environmental Consumption Externality"](#). The letters refer to the area in [Figure 9.9 "Tariff Effects and Environmental Costs"](#).

Table 9.15 Welfare Effects of a Domestic Consumption Tax with a Negative Environmental Consumption Externality

	Importing Country
Consumer Surplus	$-(c + d + e + f)$
Producer Surplus	0
Govt. Revenue	$+ c + d + e$
Pollution Effect	$+ k$
National Welfare	$k - f$

Consumers of gasoline lose from the higher price by the area $(c + d + e + f)$. Domestic producers are unaffected because their price does not change. The government also collects tax revenue, given by $(c + d + e)$, which is the product of the consumption tax ($t = P_1 - P_2$) and the level of consumption (D_1). The net total efficiency losses from trade are given by the area (f) . However, the presence of the environmental consumption externality means there is an additional benefit caused by the reduced pollution. This benefit is represented by the area k .

The net national welfare effect of the tariff is given by the summation of all effects, $(k - f)$. Since the curves are drawn such that the area k is larger than f , a consumption tax results in an improvement in national welfare in this example. More generally, we can only say that a consumption tax may result in an increase in national welfare since it will depend on the shapes of the curves and the size of the tax.

A Comparison: Trade Policy versus Domestic Policy

More interesting is the comparison between the welfare effects of a tariff and those of a consumption tax. Since the two policies are set at identical levels, it is easy to compare the effects. The distributional effects—that is, who wins and who loses—are slightly different in the two cases. First, the effects on consumers are the same since both policies raise the price to the same level. However, domestic producers suffer a loss in producer surplus with a tariff, whereas they are unaffected by the consumption tax. To some, this may look like a bad effect since domestic production of the polluting good is not reduced with the consumption tax. However, it is the net effect that matters. Next, the government collects more revenue with the domestic tax than with the tariff since both taxes are set at the same rate and consumption is greater than imports. Finally, the environmental effect is the same for both since consumption is reduced to the same level.

The net welfare effect of the consumption tax ($NW^C = k - f$) clearly must exceed the net welfare effect of a tariff ($NW^T = k - d - f$)—that is, $NW^C > NW^T$. The reason is that the tariff incurs two separate costs on society to receive the environmental benefit, whereas the consumption tax incurs only one cost for the same benefit. Specifically, the tariff causes a loss in both consumption and production efficiency (d and f), while the consumption tax only causes a consumption efficiency loss (f). For this reason, we say it is more efficient (i.e., less costly) to use a domestic consumption tax to correct for a negative consumption externality such as pollution than to use a trade policy, even though the trade policy may improve national welfare.

A Source of Controversy

For many environmental advocates, trade liberalization, or globalization more generally, clearly has the potential to cause environmental damage to many ecosystems. Concerns include pollution from industrial production, pollution from consumption, clear-cutting of tropical forests, extinction of plant and animal species, and global warming, among others. Although only one type of environmental problem is addressed above, the principles of the theory of the second best will generally apply to all these concerns.

The analysis above accepts the possibility that consumption causes pollution and that pollution is bad for society. The model shows that under these assumptions, a trade policy can potentially be used to improve environmental outcomes and can even be in society's overall interest. However, a trade policy is not the most efficient means to achieve the end. Instead, resources will be better allocated if a domestic policy, such as a consumption tax, is used instead. Since the domestic policy attacks the distortion most directly, it minimizes the economic cost. For this reason, a properly chosen consumption tax will always do better than any tariff.

With respect to other types of environmental problems, a similar conclusion can be reached. The best way to correct for most pollution and other environmental problems will be to use a domestic policy intervention such as a production tax, consumption tax, factor-use tax, or another type of domestic regulation. Trade policies, although potentially beneficial, are not the most efficient policy instruments to use.

It is worth emphasizing that the goal of most economic analysis should in many instances be aligned with the goal of environmentalists. It is the extraction and use of natural resources that contributes to environmental damage. At the same time, it is the extraction and use of natural resources that is necessary to produce the goods and services needed to raise human standards of living to acceptable levels. Thus, if we minimize the use of resources to produce a particular level of output, we can achieve both the economist's goal of maximizing efficiency and the environmentalist's goal of minimizing damage to the environment.

Understanding the WTO's Position on Trade and the Environment

In October 1999, the WTO Committee on Trade and Environment, a committee set up during the Uruguay Round to consider the linkages between these two concerns, issued its Trade and Environment report. The report argued that "there is no basis for the sweeping generalizations that are often heard in the public debate, arguing that trade is either good for the environment, or bad for the environment. The real

world linkages are a little bit of both.” World Trade Organization, “Trade liberalization reinforces the need for environmental cooperation,” press release, October 8, 1999, http://www.wto.org/english/news_e/pres99_e/pr140_e.htm.

Some of the main findings of the report are listed here with a brief explanation of how these statements relate to the theory of the second best.

Most environmental problems result from polluting production processes, certain kinds of consumption, and the disposal of waste products—trade as such is rarely the root cause of environmental degradation, except for the pollution associated with transportation of goods. World Trade Organization, “Trade liberalization reinforces the need for environmental cooperation,” press release, October 8, 1999, http://www.wto.org/english/news_e/pres99_e/pr140_e.htm.

This statement relates to the theory of the second best by highlighting that the root cause of most environmental problems is the production, consumption, and disposal processes rather than trade. The one exception is pollution caused by ships, trucks, trains, and planes transporting goods across borders, but this is a relatively minor source of global pollution. Recall that first-best solutions are those that attack the root cause of a problem most directly.

Environmental degradation occurs because producers and consumers are not always required to pay for the costs of their actions. World Trade Organization, “Trade liberalization reinforces the need for environmental cooperation,” press release, October 8, 1999, http://www.wto.org/english/news_e/pres99_e/pr140_e.htm.

This statement means that environmental problems are a negative externality in either production or consumption. If producers and consumers had to pay for the environmental effects of their actions, that would mean there is a market for pollution. In a market, the costs and benefits are internalized in the decision-making process. However, in the absence of a market, producer and consumer effects occur “external” to the market, hence the term “externality.”

However, this statement exaggerates one thing if it suggests that environmental degradation would not occur if consumers and producers were required to pay for their actions. In actuality, if a market for pollution existed, producers and consumers would continue to pollute up to the level where the costs of additional pollution exceeded the benefits. This undoubtedly would occur at some positive level of pollution and environmental degradation. As demonstrated in every environmental economics course, the socially optimal level of pollution is not zero.

Environmental degradation is sometimes accentuated by policy failures, including subsidies to polluting and resource-degrading activities—such as subsidies to agriculture, fishing and energy. World Trade Organization, “Trade liberalization reinforces the need for environmental cooperation,” press release, October 8, 1999, http://www.wto.org/english/news_e/pres99_e/pr140_e.htm.

This statement points out that many environmental problems are made worse by government interventions designed to serve some other purpose. For example, subsidies to agricultural production, designed to support the income of farmers, can have the unintended effect of encouraging the greater use of pesticides and fertilizers, thus causing a negative environmental effect. Again, this suggests that the source of environmental problems is typically not international trade.

Trade would unambiguously raise welfare if proper environmental policies were in place. World Trade Organization, “Trade liberalization reinforces the need for environmental cooperation,” press release, October 8, 1999, http://www.wto.org/english/news_e/pres99_e/pr140_e.htm.

Here, “proper environmental policies” means first-best domestic policies targeted at the environmental market failures and the elimination of other domestic policies with the unintended environmental consequences mentioned above. If these domestic policies were in place, then free trade would unambiguously be the first-best trade policy.

Trade barriers generally make for poor environmental policy. World Trade Organization, “Trade liberalization reinforces the need for environmental cooperation,” press release, October 8, 1999, http://www.wto.org/english/news_e/pres99_e/pr140_e.htm.

Why? Because of the theory of the second best. It is generally better to correct environmental externality problems using first-best domestic taxes, subsidies, or regulations than to use second-best trade policies. Thus, although trade policies *can* have favorable environmental effects, governments can achieve the same results more efficiently—that is, at a lower resource cost—by using domestic policies instead.

This is one of the strongest arguments for excluding an explicit link between environment and trade in the WTO accords and more generally in free trade area agreements. Linking the two together in a trade agreement will surely lead to the avoidance of trade liberalization in some sectors in order to secure a favorable

environmental outcome, and this will mean using trade barriers as a tool for environmental policy.

So what can or should be done? First, it is important to recognize that the WTO agreement does not prohibit countries from setting their own environmental standards. What the WTO accord does require is that countries apply most-favored nation (MFN) and national treatment in their application of environmental laws. For example, the WTO agreement does not allow a country to set one environmental standard with respect to goods imported from Argentina and another for goods from Mexico. This would violate MFN. Also, the WTO agreement would not allow a country to treat imported goods differently from goods produced at home. This would violate national treatment.

In fact, most of the WTO dispute settlement rulings (if not all) identified by environmental groups as forcing countries to change (and make more lenient) their environmental laws were not decisions to force a particular environmental standard on countries. Instead, they were decisions to enforce MFN or national treatment. Countries could have complied with any of these rulings by strengthening environmental regulations just as long as they did not discriminate internationally in their application.

Lastly, countries are not prohibited by the WTO agreement from negotiating and implementing international environmental agreements. A prime example is the Kyoto Protocol. This agreement would require signatory countries to reduce their domestic carbon emissions to agreed-on levels within a specified period of time in order to mitigate an important source of global warming. The mechanism used to reduce emissions in this case would be purely domestic policies implemented simultaneously by all signatory countries. As such, this would more likely be a first-best method to correct for global warming and would dominate any type of trade policy to solve the same problem.

One Final Issue: Measurement Problems

In the previous analysis, we assumed the environmental costs of consumption are measurable in dollar terms. However, obtaining these costs is not a simple exercise since there is no market in which pollution is traded. It may be relatively easy to measure the average amount of pollutants (carbon dioxide, sulfur, etc.) caused by each gallon of gasoline consumed, but translating that into a dollar equivalent is not a simple task. Ideally, we would want to know how much people would be willing to pay to prevent the pollution caused by each gallon of consumed gasoline. Environmental economists have tried to measure these types of costs using “contingent valuation” techniques. However, these methods are still in their

infancy in terms of providing an accurate and believable measure of environmental cost.

Without good information concerning environmental costs, it becomes almost impossible to set policies appropriately. Although welfare-improving tariffs and domestic policies can raise national welfare, they must be set at correct levels to achieve a welfare-enhancing effect. To obtain the optimal levels requires accurate information about both the economic costs and the benefits of price changes and the environmental effects as well. Without good information, it becomes more likely policies will not achieve the intended effect.

An alternative method to measure costs is for the government to require permits that allow one to pollute. If these permits were tradable, the market price of a permit would provide a reasonable estimate of the pollution cost to society. In essence, this creates a market for pollution. These programs have been applied to control industrial pollutants but have not been used in consumer markets. In addition, to most noneconomists, providing permits that allow pollution seems anathema. However, because these programs attempt to correct for problems related to the measurement of environmental costs, they may be even more efficient even than using domestic taxes.

In the end, we must recognize that our theoretical analysis can only suggest the possibility that trade liberalization will make a country worse off due to increases in pollution. The model shows that this is logically possible. However, the model also shows it is logically possible for trade liberalization to raise national welfare despite increases in pollution. It then becomes an empirical question of what the effect of trade liberalization will be. For this reason, many environmental groups, such as Sierra Club, have proposed that an environmental impact statement (EIS) be prepared for every trade agreement. An EIS would assess the environmental costs of the agreement and thereby make environmental concerns a criterion in the decision process. Presumably, these studies could prevent environmentally unfriendly trade agreements from being ratified.

Many proponents of freer trade have objected to this proposal. Jagdish Bhagwati, for one, in his book *In Defense of Globalization*, suggests that the ability to measure the environmental costs may be as difficult as, or perhaps more difficult than, measuring the economic effects of a trade agreement.

KEY TAKEAWAYS

- Environmental problems generally correspond to negative production or consumption externalities. Thus these issues represent market imperfections.
- This section presents a model in which domestic consumption of an import good causes environmental pollution (e.g., gasoline consumption). This is the case of a negative consumption externality.
- The model is used to show that trade liberalization may cause a reduction in national welfare if the additional pollution caused by increased consumption is greater than the efficiency benefits that arise from freer trade. Thus concerns that trade liberalization may cause environmental damage are consistent with economic theory.
- However, the theory of the second best suggests that when market imperfections exist, trade policy corrections may be second-best, not first-best, policy choices.
- Both an import tariff and a domestic consumption tax will reduce domestic consumption of the import good and lead to a reduction in pollution. However, the domestic consumption tax achieves the result at a lower economic cost than the import tariff. Thus we say that the domestic consumption tax is a first-best policy, while the import tariff is a second-best policy.
- The previous result corresponds to the general theory of the second best, which says that the first-best policy will be the policy that targets a distortion or imperfection most directly. In most cases, a domestic policy will be better than a trade policy. In this example, a domestic consumption tax is clearly superior to a trade policy.

EXERCISE

1. Consider a perfectly competitive market for steel in a small exporting country. Suppose that steel production causes local air and water pollution. Assume that the larger is steel output and the higher is the social cost of pollution; thus, steel production creates a negative externality.
 - a. Explain what type of trade tax or subsidy policy could be used to reduce the negative effects of pollution. Demonstrate the welfare effects using a partial equilibrium diagram. Assume that your policy reduces social costs by R dollars. Under what condition would the policy raise national welfare?
 - b. Explain what type of purely domestic policy could be used to reduce the pollution. Use a partial equilibrium diagram to demonstrate the welfare effects of this policy. Again assume that your policy reduces social costs by R dollars.
 - c. Explain why the purely domestic policy may be superior to the trade policy.

9.10 Economic Integration: Free Trade Areas, Trade Creation, and Trade Diversion

LEARNING OBJECTIVES

1. Distinguish the different types of economic integration.
2. Learn the effects of trade creation and trade diversion.
3. Understand how free trade area formation can make a country worse off in terms of the theory of the second best.

For a variety of reasons, it often makes sense for nations to coordinate their economic policies. Coordination can generate benefits that are not possible otherwise. A clear example of this is shown in the discussion of trade wars among large countries in [Chapter 7 "Trade Policy Effects with Perfectly Competitive Markets"](#), [Section 7.9 "Retaliation and Trade Wars"](#). There it is shown that if countries cooperate and set zero tariffs against each other, then both countries are likely to benefit relative to the case when both countries attempt to secure short-term advantages by setting optimal tariffs. This is just one advantage of cooperation. Benefits may also accrue to countries that liberalize labor and capital movements across borders, that coordinate fiscal policies and resource allocation toward agriculture and other sectors, and that coordinate their monetary policies.

Any type of arrangement in which countries agree to coordinate their trade, fiscal, or monetary policies is referred to as **economic integration**¹⁴. There are many different degrees of integration.

Preferential Trade Agreement

A preferential trade agreement (PTA) is perhaps the weakest form of economic integration. In a PTA, countries would offer tariff reductions, though perhaps not eliminations, to a set of partner countries in some product categories. Higher tariffs, perhaps nondiscriminatory tariffs, would remain in all other product categories. This type of trade agreement is not allowed among World Trade Organization (WTO) members, who are obligated to grant most-favored nation (MFN) status to all other WTO members. Under the MFN rule, countries agree not to discriminate against other WTO member countries. Thus, if a country's low tariff on bicycle imports, for example, is 5 percent, then it must charge 5 percent on imports from all other WTO members. Discrimination or preferential treatment for some countries is not allowed. The country is free to charge a higher tariff on imports

¹⁴. Any type of arrangement in which countries agree to coordinate their trade, fiscal, or monetary policies.

from non-WTO members, however. In 1998, the United States proposed legislation to eliminate tariffs on imports from the nations in sub-Saharan Africa. This action represents a unilateral preferential trade agreement since tariffs would be reduced in one direction but not the other. (Note that a PTA is also used more generally to describe all types of economic integration since they all incorporate some degree of “preferred” treatment.)

Free Trade Area

A **free trade area (FTA)**¹⁵ occurs when a group of countries agrees to eliminate tariffs among themselves but maintain their own external tariff on imports from the rest of the world. The North American Free Trade Agreement (NAFTA) is an example of an FTA. When NAFTA is fully implemented, tariffs of automobile imports between the United States and Mexico will be zero. However, Mexico may continue to set a different tariff than the United States on automobile imports from non-NAFTA countries. Because of the different external tariffs, FTAs generally develop elaborate “rules of origin.” These rules are designed to prevent goods from being imported into the FTA member country with the lowest tariff and then transshipped to the country with higher tariffs. Of the thousands of pages of text that make up NAFTA, most of them describe rules of origin.

Customs Union

A customs union occurs when a group of countries agrees to eliminate tariffs among themselves and set a common external tariff on imports from the rest of the world. The European Union (EU) represents such an arrangement. A customs union avoids the problem of developing complicated rules of origin but introduces the problem of policy coordination. With a customs union, all member countries must be able to agree on tariff rates across many different import industries.

Common Market

A common market establishes free trade in goods and services, sets common external tariffs among members, and also allows for the free mobility of capital and labor across countries. The EU was established as a common market by the Treaty of Rome in 1957, although it took a long time for the transition to take place. Today, EU citizens have a common passport, can work in any EU member country, and can invest throughout the union without restriction.

15. A situation in which a group of countries agrees to eliminate tariffs among themselves but maintain their own external tariff on imports from the rest of the world.

Economic Union

An economic union typically will maintain free trade in goods and services, set common external tariffs among members, allow the free mobility of capital and labor, and also relegate some fiscal spending responsibilities to a supranational agency. The EU's Common Agriculture Policy (CAP) is an example of a type of fiscal coordination indicative of an economic union.

Monetary Union

A monetary union establishes a common currency among a group of countries. This involves the formation of a central monetary authority that will determine monetary policy for the entire group. The Maastricht treaty, signed by EU members in 1992, proposed the implementation of a single European currency (the Euro) by 1999.

Perhaps the best example of an economic and monetary union is the United States. Each U.S. state has its own government that sets policies and laws for its own residents. However, each state cedes control, to some extent, over foreign policy, agricultural policy, welfare policy, and monetary policy to the federal government. Goods, services, labor, and capital can all move freely, without restrictions among the U.S. states, and the nation sets a common external trade policy.

Multilateralism versus Regionalism

In the post-World War II period, many nations pursued the objective of trade liberalization. One device used to achieve this was the General Agreement on Tariffs and Trade (GATT) and its successor, the WTO. Although the GATT began with less than 50 member countries, the WTO now claims 153 members as of 2010. Since GATT and WTO agreements commit all member nations to reduce trade barriers simultaneously, the agreements are sometimes referred to as a *multilateral* approach to trade liberalization.

An alternative method used by many countries to achieve trade liberalization includes the formation of preferential trade arrangements, free trade areas, customs unions, and common markets. Since many of these agreements involve geographically contiguous countries, these methods are sometimes referred to as a *regional* approach to trade liberalization.

The key question of interest concerning the formation of preferential trade arrangements is whether these arrangements are a good thing. If so, under what conditions? If not, why not?

One reason supporters of free trade may support regional trade arrangements is because they are seen to represent movements toward free trade. Indeed, Section 24 of the original GATT allows signatory countries to form free trade agreements and customs unions despite the fact that preferential agreements violate the principle of nondiscrimination. When a free trade area or customs union is formed between two or more WTO member countries, they agree to lower their tariffs to zero between each other but will maintain their tariffs against other WTO countries. Thus the free trade area is a discriminatory policy. Presumably, the reason these agreements are tolerated within the WTO is because they represent significant commitments to free trade, which is another fundamental goal of the WTO.

However, there is also some concern among economists that regional trade agreements may make it more difficult, rather than easier, to achieve the ultimate objective of global free trade.

The fear is that although regional trade agreements will liberalize trade among their member countries, the arrangements may also increase incentives to raise protectionist trade barriers against countries outside the area. The logic here is that the larger the regional trade area relative to the size of the world market, the larger will be that region's market power in trade. The more market power, the higher would be the region's optimal tariffs and export taxes. Thus the regional approach to trade liberalization could lead to the formation of large "trade blocs" that trade freely among members but choke off trade with the rest of the world. For this reason, some economists have argued that the multilateral approach to trade liberalization, represented by the trade liberalization agreements in successive WTO rounds, is more likely to achieve global free trade than the regional or preferential approach.

Much has been written on this subject recently. Here we have merely scratched the surface.

In what follows, we present the economic argument regarding trade diversion and trade creation. These concepts are used to distinguish between the effects of free trade area or customs union formation that may be beneficial and those that are detrimental. As mentioned, preferential trade arrangements are often supported because they represent a movement in the direction of free trade. If free trade is economically the most efficient policy, it would seem to follow that any movement toward free trade should be beneficial in terms of economic efficiency. It turns out that this conclusion is wrong. Even if free trade is most efficient, it is not true that a step in that direction necessarily raises economic efficiency. Whether a preferential trade arrangement raises a country's welfare and raises economic efficiency

depends on the extent to which the arrangement causes trade diversion versus trade creation.

Trade Creation and Trade Diversion

In this section, we present an analysis of trade diversion and trade creation. The analysis uses a partial equilibrium framework, which means that we consider the effects of preferential trade liberalization with respect to a representative industry. Later in the section we consider how the results from the representative industry cases can be extended to consider trade liberalization that covers all trade sectors.

We assume in each case that there are three countries in the world: Countries A, B, and C. Each country has supply and demand for a homogeneous good in the representative industry. Countries A and B will form a free trade area. (Note that trade diversion and creation can occur regardless of whether a preferential trade agreement, a free trade area, or a customs union is formed. For convenience, we'll refer to the arrangement as a free trade area [FTA].) The attention in this analysis will be on Country A, one of the two FTA members. We'll assume that Country A is a small country in international markets, which means that it takes international prices as given. Countries B and C are assumed to be large countries (or regions). Thus Country A can export or import as much of a product as desired with Countries B and C at whatever price prevails in those markets.

We assume that if Country A were trading freely with either B or C, it would wish to import the product in question. However, Country A initially is assumed *not* to be trading freely. Instead, the country will have an MFN-specific tariff (i.e., the same tariff against both countries) applied on imports from both Countries B and C.

In each case below, we will first describe an initial tariff-ridden equilibrium. Then, we will calculate the price and welfare effects that would occur in this market if Countries A and B form an FTA. When the FTA is formed, Country A maintains the same tariff against Country C, the non-FTA country.

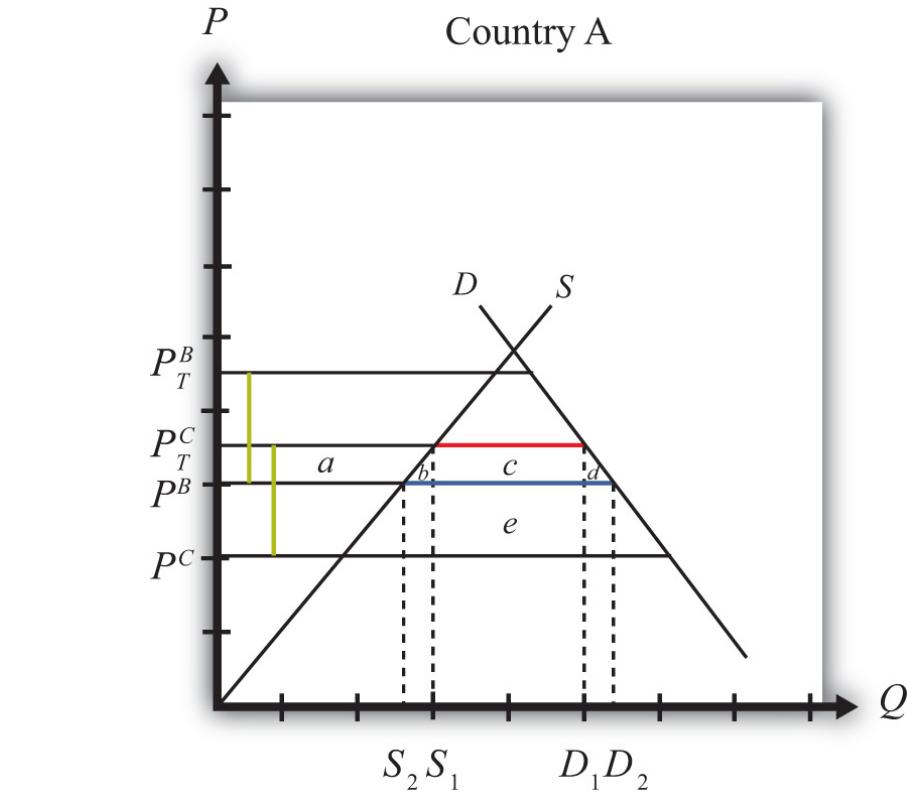
Trade Diversions

In general, a **trade diversion**¹⁶ means that a free trade area diverts trade away from a more-efficient supplier outside the FTA and toward a less-efficient supplier within the FTA. In some cases, trade diversion will reduce a country's national welfare, but in some cases national welfare could improve despite the trade diversion. We present both cases below.

16. A situation in which a free trade area diverts trade away from a more-efficient supplier outside the FTA toward a less-efficient supplier within the FTA.

Figure 9.10 "Harmful Trade Diversion" depicts the case in which trade diversion is harmful to a country that joins an FTA. The graph shows the supply and demand curves for Country A. P^B and P^C represent the free trade supply prices of the good from Countries B and C, respectively. Note that Country C is assumed to be capable of supplying the product at a lower price than Country B. (Note that in order for this to be possible, Country B must have tariffs or other trade restrictions on imports from Country C, or else all of B's market would be supplied by C.)

Figure 9.10 *Harmful Trade Diversion*



We assume that A has a specific tariff $t^B = t^C = t^*$ set on imports from both Countries B and C. The tariff raises the domestic supply prices to P_T^B and P_T^C , respectively. The size of the tariff is denoted by the green dotted lines in [Figure 9.10 "Harmful Trade Diversion"](#), which show that $t^* = P_T^B - P^B = P_T^C - P^C$.

Since, with the tariff, the product is cheaper from Country C, Country A will import the product from Country C and will not trade initially with Country B. Imports are given by the red line, or by the distance $D^1 - S^1$. Initial tariff revenue is given by the area $(c + e)$, the tariff rate multiplied by the quantity imported.

Next, assume Countries A and B form an FTA and A eliminates the tariff on imports from Country B. Now, $t^B = 0$, but t^C remains at t^* . The domestic prices on goods from Countries B and C are now P^B and P_T^C , respectively. Since $P^B < P_T^C$, Country A would import all the product from Country B after the FTA and would import nothing from Country C. At the lower domestic price, P^B , imports would rise to $D^2 - S^2$, denoted by the blue line. Also, since the nondistorted (i.e., free trade) price in Country C is less than the price in Country B, trade is said to be *diverted* from a more-efficient supplier to a less-efficient supplier.

The welfare effects are summarized in [Table 9.16 "Welfare Effects of Free Trade Area Formation: Trade Diversion Cases"](#).

Table 9.16 Welfare Effects of Free Trade Area Formation: Trade Diversion Cases

	Country A
Consumer Surplus	+ (a + b + c + d)
Producer Surplus	- a
Govt. Revenue	- (c + e)
National Welfare	+ (b + d) - e

Free trade area effects on Country A's consumers. Consumers of the product in the importing country benefit from the free trade area. The reduction in the domestic price of both the imported goods and the domestic substitutes raises consumer surplus in the market. Refer to [Table 9.16 "Welfare Effects of Free Trade Area Formation: Trade Diversion Cases"](#) and [Figure 9.10 "Harmful Trade Diversion"](#) to see how the magnitude of the change in consumer surplus is represented.

Free trade area effects on Country A's producers. Producers in the importing country suffer losses as a result of the free trade area. The decrease in the price of their product on the domestic market reduces producer surplus in the industry. The price decrease also induces a decrease in the output of existing firms (and perhaps some firms will shut down), a decrease in employment, and a decrease in profit, payments, or both to fixed costs. Refer to [Table 9.16 "Welfare Effects of Free Trade Area Formation: Trade Diversion Cases"](#) and [Figure 9.10 "Harmful Trade Diversion"](#) to see how the magnitude of the change in producer surplus is represented.

Free trade area effects on Country A's government. The government loses all the tariff revenue that had been collected on imports of the product. This reduces government revenue, which may in turn reduce government spending or transfers

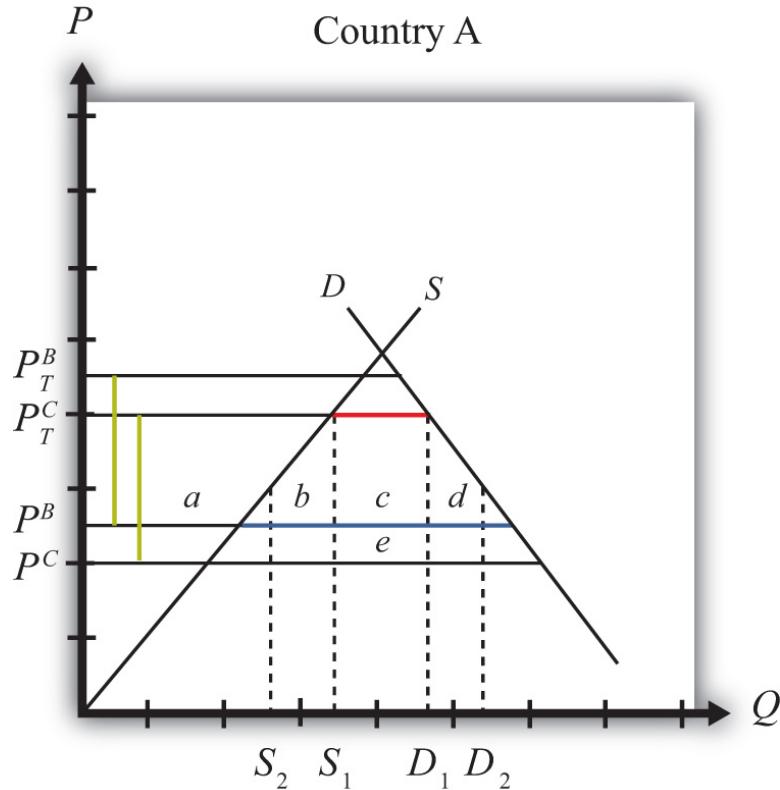
or raise government debt. Who loses depends on how the adjustment is made. Refer to [Table 9.16 "Welfare Effects of Free Trade Area Formation: Trade Diversion Cases"](#) and [Figure 9.10 "Harmful Trade Diversion"](#) to see how the magnitude of the tariff revenue is represented.

Free trade area effects on Country A's national welfare. The aggregate welfare effect for the country is found by summing the gains and losses to consumers, producers, and the government. The net effect consists of three components: a positive production efficiency gain (b), a positive consumption efficiency gain (d), and a negative tariff revenue loss (e). Notice that not all the tariff revenue loss ($c + e$) is represented in the loss to the nation. That's because some of the total losses (area c) are, in effect, transferred to consumers. Refer to [Table 9.16 "Welfare Effects of Free Trade Area Formation: Trade Diversion Cases"](#) and [Figure 9.10 "Harmful Trade Diversion"](#) to see how the magnitude of the change in national welfare is represented.

Because there are both positive and negative elements, the net national welfare effect can be either positive or negative. [Figure 9.10 "Harmful Trade Diversion"](#) depicts the case in which the FTA causes a reduction in national welfare. Visually, it seems obvious that area e is larger than the sum of a and b . Thus, under these conditions, the FTA with trade diversion would cause national welfare to fall.

If conditions were different, however, the national welfare change could be positive. Consider [Figure 9.11 "Beneficial Trade Diversion"](#). This diagram differs from [Figure 9.10 "Harmful Trade Diversion"](#) only in that the free trade supply price offered by Country B, P^B , is lower and closer to Country C's free trade supply price, P^C . The description earlier concerning the pre- and post-FTA equilibria remains the same, and trade diversion still occurs. The welfare effects remain the same in direction but differ in magnitude. Notice that the consumer surplus gain is now larger because the drop in the domestic price is larger. Also notice that the net national welfare effect, $(b + d - e)$, visually appears positive. This shows that in some cases, formation of an FTA that causes a trade diversion may have a positive net national welfare effect. Thus a trade diversion may be, but is not necessarily, welfare reducing.

Figure 9.11 *Beneficial Trade Diversion*



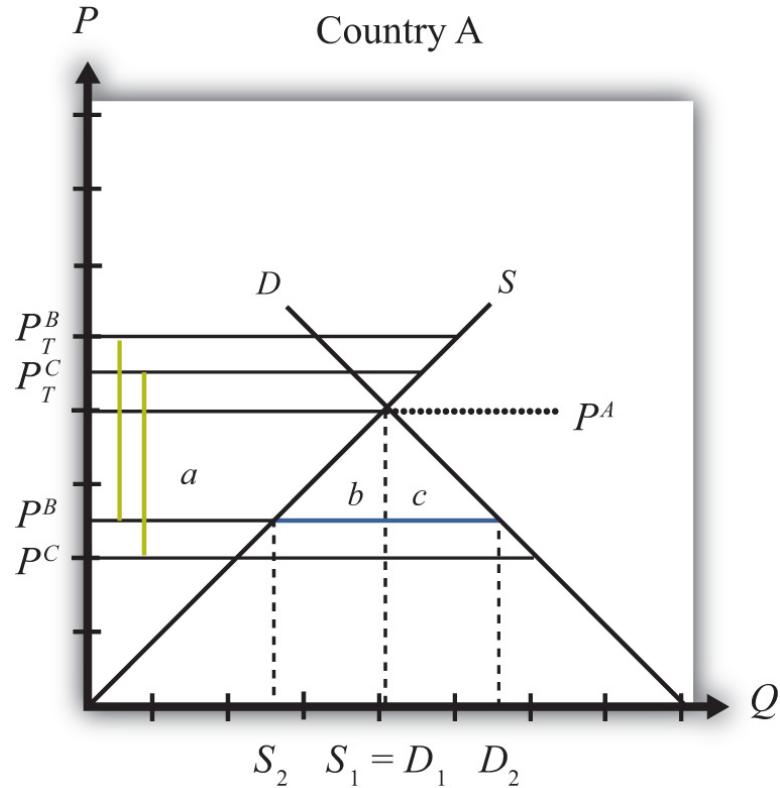
Generally speaking, the larger the difference between the nondistorted prices in the FTA partner country and in the rest of the world, the more likely it is that trade diversion will reduce national welfare.

Trade Creation

In general, **trade creation**¹⁷ means that a free trade area creates trade that would not have existed otherwise. As a result, supply occurs from a more-efficient producer of the product. In all cases, trade creation will raise a country's national welfare.

Figure 9.12 "Trade Creation" depicts a case of trade creation. The graph shows the supply and demand curves for Country A. P^B and P^C represent the free trade supply prices of the good from Countries B and C, respectively. Note that Country C is assumed to be capable of supplying the product at a lower price than Country B. (Note that in order for this to be possible, Country B must have tariffs or other trade restrictions on imports from Country C, or else all of B's market would be supplied by C.)

17. A situation in which a free trade area creates trade that would not have existed otherwise.

Figure 9.12 *Trade Creation*


We assume that A has a specific tariff, $t^B = t^C = t^*$, set on imports from both Countries B and C. The tariff raises the domestic supply prices to P_T^B and P_T^C , respectively. The size of the tariff is denoted by the green dotted lines in Figure 9.12 "Trade Creation", which show that $t^* = P_T^B - P^B = P_T^C - P^C$.

Since, with the tariffs, the autarky price in Country A, labeled P^A in Figure 9.12 "Trade Creation", is less than the tariff-ridden prices P_T^B and P_T^C , the product will not be imported. Instead, Country A will supply its own domestic demand at $S^1 = D^1$. In this case, the original tariffs are prohibitive.

Next, assume Countries A and B form an FTA and A eliminates the tariff on imports from Country B. Now $t^B = 0$, but t^C remains at t^* . The domestic prices on goods from Countries B and C are now P^B and P_T^C , respectively. Since $P^B < P^A$, Country A would now import the product from Country B after the FTA. At the lower domestic price P^B , imports would rise to the blue line distance, or $D_2 - S_2$. Since trade now occurs with the FTA and it did not occur before, trade is said to be *created*.

The welfare effects are summarized in [Table 9.17 "Welfare Effects of Free Trade Area Formation: Trade Creation Case"](#).

Table 9.17 Welfare Effects of Free Trade Area Formation: Trade Creation Case

	Country A
Consumer Surplus	+ (a + b + c)
Producer Surplus	- a
Govt. Revenue	0
National Welfare	+ (b + c)

Free trade area effects on Country A's consumers. Consumers of the product in the importing country benefit from the free trade area. The reduction in the domestic price of both imported goods and the domestic substitutes raises consumer surplus in the market. Refer to [Table 9.17 "Welfare Effects of Free Trade Area Formation: Trade Creation Case"](#) and [Figure 9.12 "Trade Creation"](#) to see how the magnitude of the change in consumer surplus is represented.

Free trade area effects on Country A's producers. Producers in the importing country suffer losses as a result of the free trade area. The decrease in the price of their product in the domestic market reduces producer surplus in the industry. The price decrease also induces a decrease in output of existing firms (and perhaps some firms will shut down), a decrease in employment, and a decrease in profit, payments, or both to fixed costs. Refer to [Table 9.17 "Welfare Effects of Free Trade Area Formation: Trade Creation Case"](#) and [Figure 9.12 "Trade Creation"](#) to see how the magnitude of the change in producer surplus is represented.

Free trade area effects on Country A's government. Since initial tariffs were prohibitive and the product was not originally imported, there was no initial tariff revenue. Thus the FTA induces no loss of revenue.

Free trade area effects on Country A's national welfare. The aggregate welfare effect for the country is found by summing the gains and losses to consumers and producers. The net effect consists of two positive components: a positive production efficiency gain (b) and a positive consumption efficiency gain (c). This means that if trade creation arises when an FTA is formed, it must result in net national welfare gains. Refer to [Table 9.17 "Welfare Effects of Free Trade Area Formation: Trade Creation Case"](#) and [Figure 9.12 "Trade Creation"](#) to see how the magnitude of the change in national welfare is represented.

Aggregate Welfare Effects of a Free Trade Area

The analysis above considers the welfare effects on participants in one particular market in one country that is entering into a free trade area. However, when a free trade area is formed, presumably many markets and multiple countries are affected, not just one. Thus, to analyze the aggregate effects of an FTA, one would need to sum up the effects across markets and across countries.

The simple way to do that is to imagine that a country entering an FTA may have some import markets in which trade creation would occur and other markets in which trade diversion would occur. The markets with trade creation would definitely generate national welfare gains, while the markets with trade diversion *may* generate national welfare losses. It is common for economists to make the following statement: “If the positive effects of trade creation are larger than the negative effects of trade diversion, then the FTA will improve national welfare.” A more succinct statement, though also somewhat less accurate, is that “if an FTA causes more trade creation than trade diversion, then the FTA is welfare improving.”

However, the converse statement is also possible—that is, “if an FTA causes more trade diversion than trade creation, then the FTA may be *welfare reducing* for a country.” This case is actually quite interesting since it suggests that a movement to free trade by a group of countries may actually reduce the national welfare of the countries involved. This means that a movement in the direction of a more-efficient free trade policy may not raise economic efficiency. Although this result may seem counterintuitive, it can easily be reconciled in terms of the theory of the second best.

Free Trade Areas and the Theory of the Second Best

One might ask, if free trade is economically the most efficient policy, how can it be that a movement to free trade by a group of countries can reduce economic efficiency? The answer is quite simple once we put the story of FTA formation into the context of the theory of the second best. Recall that the theory of the second best suggested that when there are distortions or imperfections in a market, then the addition of another distortion (like a trade policy) could actually raise welfare or economic efficiency. In the case of an FTA, the policy change is the removal of trade barriers rather than the addition of a new trade policy. However, the second-best theory works much the same in reverse.

Before a country enters an FTA, it has policy-imposed distortions already in place in the form of tariff barriers applied on imports of goods. This means that the initial

equilibrium can be characterized as a second-best equilibrium. When the FTA is formed, some of these distortions are removed—that is, the tariffs applied to one's FTA partners. However, other distortions remain—that is, tariffs applied against the nonmember countries. If the partial tariff removal substantially raises the negative effects caused by the remaining tariff barriers with the non-FTA countries, then the efficiency improvements caused by free trade within the FTA could be outweighed by the negative welfare effects caused by the remaining barriers outside the FTA, and national welfare could fall.

This is, in essence, what happens in the case of trade diversion. Trade diversion occurs when an FTA shifts imports from a more-efficient supplier to a less-efficient supplier, which by itself causes a reduction in national welfare. Although the economy also benefits through the elimination of the domestic distortions, if these benefits are smaller than the supplier efficiency loss, then national welfare falls. In general, the only way to assure that trade liberalization will lead to efficiency improvements is if a country removes its trade barriers against all countries.

KEY TAKEAWAYS

- Countries can integrate by reducing barriers to trade under multilateral arrangements like the WTO or by entering into regional arrangements, including preferential trade agreements, free trade agreements, customs unions, common markets, or monetary unions.
- The formation of a free trade area can lead to trade creation or trade diversion.
- Trade creation involves new trade that would not exist without the FTA and is always beneficial for the countries in terms of national welfare.
- Trade diversion involves the shifting of trade away from one country toward one's free trade partner and is sometimes detrimental to the countries in terms of national welfare.
- Losses caused by trade diversion can be understood in terms of the theory of the second best; because one market distortion remains when another is removed, welfare can fall.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. An arrangement in which a group of countries agrees to eliminate tariffs among themselves but maintain their own external tariff on imports from the rest of the world.
 - b. The term used to describe a change in the pattern of trade in response to trade liberalization in which a country begins to import from a less-efficient supplier.
 - c. The term used to describe a change in the pattern of trade in response to trade liberalization in which a country begins to import from a more-efficient supplier.

Chapter 10

Political Economy and International Trade

Trade policy analysis is often conducted from the implicit vantage point of a benevolent dictator poised to choose the best policies for a country. However, decisions about which policies to apply are rarely made by a sovereign but instead are usually made via a democratic political process. Whenever we consider how the political process affects economic decision making, we call it political economy.

The political economy aspects of trade policymaking are studied briefly in this chapter. Most important is how the concentrations and dispersion of the costs and benefits of trade policies tend to affect the decisions.

10.1 Chapter Overview

LEARNING OBJECTIVE

1. Understand the motivations of a government in determining the policies that affect international trade.

In most economic models, it is assumed that consumers maximize utility, firms maximize profit, and governments maximize national welfare. Although one can reasonably object to any one of these assumptions, perhaps the one least likely to hold is the assumption about a government's behavior. Governments are rarely comprised of a solitary decision maker whose primary interest is the maximum well-being of the nation's constituents. Such a person, if he or she existed, could be labeled a "benevolent dictator." Although historically some nations have been ruled almost single-handedly by dictators, most dictators could hardly be called benevolent.

The assumption that governments behave as if they had a benevolent dictator may have developed out of the philosophical traditions of utilitarianism. Utilitarianism, whose roots date to writings by Jeremy Bentham in the early 1800s, suggests that the objective of society should be to produce the greatest good for the greatest number. The objective of individuals is to obtain utility (happiness, satisfaction, well-being, etc.). In economic analysis, we presume that individuals obtain all their utility from the consumption of goods and services, and this motivates the behavioral assumption that consumers maximize utility. The assumption that firms maximize profit is based on the same logic. Profit affects the income of firm owners. The greater one's income, the greater will be one's consumption possibilities and thus the higher will be one's utility. Thus profit is merely a means to an end, the end being greater utility. It is not unreasonable, then, that if the objective of individuals and firms is maximum utility, then the objective of a government might be to maximize utility for everyone.

But even if governments do not seek to maximize national welfare, it is still a valid exercise to investigate which policies would lead to maximum utility. Indeed, most of the analysis of trade policies does just this. Policy analysis identifies the differential welfare effects of various policies and points out which of these will lead to the greatest overall utility or welfare.

If one prescribes policies that also maximize national welfare, then one is making the value judgment that maximum national welfare is the appropriate goal for a government. If one presumes that governments do indeed seek to maximize national welfare, then the task is to explain why the choices that governments make are explainable as the outcome of a national welfare maximization exercise. An alternative approach is to consider other reasons for the choices made by governments. This is essentially the task of political economy models.

Political economy¹ is a term that reflects the interaction between the economic system and the political system. Many traditional models of the economy make simplifying assumptions about the behavior of governments. Keeping the model simple is one reason for the assumption of a benevolent dictator. Political economy models attempt to explain more carefully the decision-making process of governments. Today, most governments can be best described as representative democracies. This means that government officials are elected, through some voting procedure, to “represent” the interests of their constituents in making government decisions.

The key issue in political economy and trade models is to explain how political features in democratic economies affect the choice of a trade policy. Among the key questions are the following:

1. Why do countries choose protection so often, especially given that economists have been emphasizing the advantages of free trade for three hundred or more years? In other words, if free trade is as good as economists say, then why do nations choose to protect?
2. In discussions of trade policies, why is so much attention seemingly given to a policy’s effects on businesses or firms and so little attention given to the effects on consumers?
3. Why do political discussions, even today, have a mercantilist spirit, wherein exports are hailed as beneficial, while imports are treated as harmful to the country?

KEY TAKEAWAYS

- Economic modelers often seem to assume that governments will choose policies to maximize the nation’s welfare.
- Instead, most government policies arise from complex decision making in a representative democracy.
- “Political economy” is a term used to describe the process of government decisions on economic policies.

1. A term used to describe the interaction between the economic system and the political system.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The name for the philosophical ideas suggesting that the purpose of society is to create the greatest good for the greatest number.
 - b. The name for a solitary leader of a country whose intention is to maximize the well-being of the nation’s constituents.
 - c. The term used to describe the interaction of the political system and the economic system.
 - d. The term used to describe a contention that exports are good and imports are bad for a country.

10.2 Some Features of a Democratic Society

LEARNING OBJECTIVE

- Understand how and why lobbying is used to influence the policy decisions of a government.

A government represents the interests of its citizens. As Abraham Lincoln said in the Gettysburg Address, a democratic government is meant to be *by* the people and *for* the people. Thus, in a **representative democracy**², government officials are entrusted to take actions that are in the interests of their constituents. Periodic elections allow citizens to vote for individuals they believe will best fulfill their interests. If elected officials do not fulfill the interests of their constituents, then those constituents eventually have a chance to vote for someone else. Thus, if elected officials are perceived as good representatives of their constituents' interests, they are likely to be reelected. If they follow their own individual agenda, and if that agenda does not match the general interests of their constituents, then they may lose a subsequent bid for reelection.

Citizens in democratic societies are traditionally granted the right to free speech. It is generally accepted that people should be allowed to voice their opinions about anything in front of others. In particular, people should be free to voice their opinions about government policies and actions without fear of reprisal. Criticisms of, as well as recommendations for, government policy actions must be allowed if a truly representative government is to operate effectively.

The Nature of Lobbying

- A political system where government officials are entrusted to take actions that are in the interests of their constituents. Periodic elections are the mechanism that makes it work.
- The activity wherein individual citizens voice their opinions to government officials about government policy actions.

We can define **lobbying**³ as the activity wherein individual citizens voice their opinions to the government officials about government policy actions. It is essentially an information transmission process. By writing letters and speaking with officials, individuals inform the government about their preferences for various policy options under consideration. We can distinguish two types of lobbying: casual lobbying and professional lobbying.

Casual lobbying occurs when a person uses his leisure time to petition or inform government officials of his point of view. Examples of casual lobbying are when people express their opinions at a town meeting or when they write letters to their Congress members. In these cases, there is no opportunity cost for the economy in

terms of lost output, although there is a cost to the individual because of the loss of leisure time. Casual lobbying, then, poses few economic costs except to the individual engaging in the activity.

Professional lobbying occurs when an individual or company is hired by someone to advocate a point of view before the government. An example is a law firm hired by the steel industry to help win an antidumping petition. In this case, the law firm will present arguments to government officials to try to affect a policy outcome. The law firm's fee will come from the extra revenue expected by the steel industry if it wins the petition. Since, in this case, the law firm is paid to provide lobbying services, there is an opportunity cost represented by the output that could have been produced had the lawyers engaged in an alternative productive activity. When lawyers spend time lobbying, they can't spend time writing software programs, designing buildings, building refrigerators, and so on. (This poses the question, What would lawyers do if they weren't lawyering?) The lawyers' actions with this type of lobbying are essentially redistributive in nature, since the lawyers' incomes will derive from the losses that will accrue to others in the event that the lobbying effort is successful. If the lobbying effort is not successful, the lawyers will still be paid, only this time the losses will accrue to the firm that hired the lawyers. For this reason, lobbying is often called **rent seeking**⁴ because the fees paid to the lobbyists come from a pool of funds (rents) that arise when the lobbying activity is successful. Another name given to professional lobbying in the economics literature is a **directly unproductive profit-seeking (DUP) activity**⁵.

Lobbying is necessary for the democratic system to work. Somehow information about preferences and desires must be transmitted from citizens to the government officials who make policy decisions. Since everyone is free to petition the government, lobbying is the way in which government officials can learn about the desires of their constituents. Those who care most about an issue will be more likely to voice their opinions. The extent of the lobbying efforts may also inform the government about the intensity of the preferences as well.

4. A term used to describe the purposeful activity of seeking ways to shift profit or "rents" toward oneself or a favored group.

5. Any activity whose primary purpose is to shift benefits toward a particular group, often by influencing policy decisions. Although a DUP activity may indirectly benefit a producer of final consumer goods or services, it does not directly contribute to that production.

KEY TAKEAWAYS

- In a representative democracy, citizens have the right to both elect their representatives and discuss policy options with their elected representatives.
- Lobbying is the process of providing information to elected officials to influence the policies that are implemented.
- A directly unproductive profit-seeking (DUP) activity is any action that by itself does not directly produce final goods and services consumed by a country's consumers.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term for a democratic system in which government agents are entrusted to take actions in the best interests of the voting public.
 - b. The term used to describe activities to petition the government for particular policies.
 - c. This type of lobbying does not incur an opportunity cost of forgone production.
 - d. This type of lobbying does incur an opportunity cost of forgone production.
 - e. The term used to describe the extra revenues earned because of successful lobbying efforts.
 - f. The term describing the purposeful effort to direct money away from others and toward oneself.
 - g. Economic activities defined by the acronym DUP.

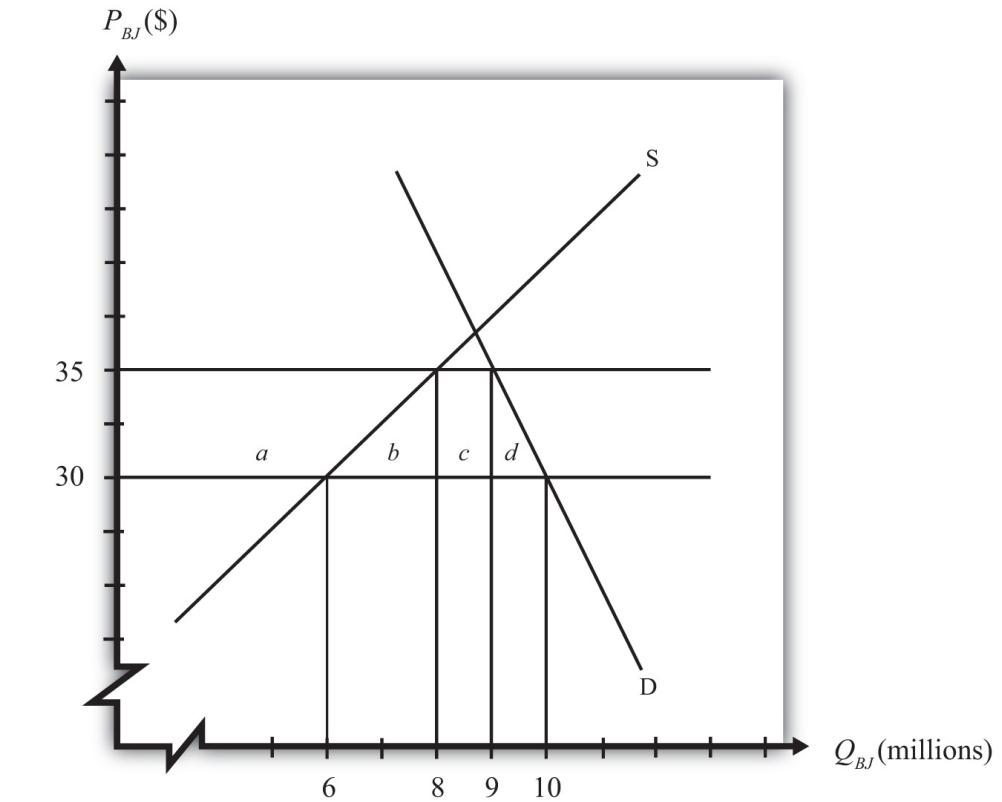
10.3 The Economic Effects of Protection: An Example

LEARNING OBJECTIVE

1. Depict numerical values for the welfare effects of a tariff by a small country.

Consider the market for blue jeans in a small importing country, depicted in [Figure 10.1 "A Market for Blue Jeans"](#). Suppose a sudden increase in the world supply of jeans causes the world market price to fall from \$35 to \$30. The price decrease causes an increase in domestic demand from nine to ten million pairs of jeans, a decrease in domestic supply from eight to six million pairs, and an increase in imports from one to four million.

Figure 10.1 A Market for Blue Jeans



Because of these market changes, suppose that the import-competing industry uses its trade union to organize a petition to the government for temporary protection. Let's imagine that the industry calls for a \$5 tariff so as to reverse the effects of the import surge. Note that this type of action is allowable to World Trade Organization (WTO) member countries under the "escape clause" or "safeguards clause."

We can use the measures of producer surplus and consumer surplus to calculate the effects of a \$5 tariff. These effects are summarized in [Table 10.1 "Welfare Effects of an Import Tariff"](#). The dollar values are calculated from the respective areas on the graph in [Figure 10.1 "A Market for Blue Jeans"](#).

Table 10.1 Welfare Effects of an Import Tariff

	Area on Graph	\$ Value
Consumer Surplus	$-(a + b + c + d)$	-\$47.5 million
Producer Surplus	$+a$	+\$35 million
Govt. Revenue	$+c$	+\$5 million
National Welfare	$-(b + d)$	-\$7.5 million

Notice that consumers lose more than the gains that accrue to the domestic producers and the government combined. This is why national welfare is shown to decrease by \$7.5 million.

In order to assess the political ramifications of this potential policy, we will make some additional assumptions. In most markets, the number of individuals that makes up the demand side of the market is much larger than the number of firms that makes up the domestic import-competing industry. Suppose, then, that the consumers in this market are made up of millions of individual households, each of which purchases, at most, one pair of jeans. Suppose the domestic blue jeans industry is made up of thirty-five separate firms.

KEY TAKEAWAY

- With quantities, prices, and the tariff rate specified, actual values for the changes in consumer and producer surplus and government revenue can be determined.

EXERCISE

1. Suppose the supply and demand curves for bottles of Coke are given by,

$$S = 10P - 7$$

$$D = 13 - 5P$$

where P is the price of Coke per bottle, D is the quantity of Coke demand (in millions of bottles), and S is the quantity of Coke supply (in millions of bottles). Suppose the free trade price of Coke is \$1.00 and that a tariff of \$0.20 is being considered by the government. If the country is a small importer calculate the following:

- a. The value of the increase in producer surplus expected due to the tariff.
- b. The value of the decrease in consumer surplus expected due to the tariff.
- c. The value of the tariff revenue expected due to the tariff.
- d. The value of the change in national welfare expected due to the tariff.

10.4 The Consumers' Lobbying Decision

LEARNING OBJECTIVE

1. Learn the lobbying implications of the widely dispersed costs of protection to consumers.

If the \$5 tariff is implemented, it will raise the price from \$30 to \$35. Consumption will fall from ten million to nine million pairs of jeans. Because of our simplifying assumption of one household per pair of jeans, one million households will decide not to purchase jeans because of the higher price. They will use the \$35 to buy something else they think is more valuable than jeans. The other nine million households will pay the extra \$5. This means that, at most, a household has to pay an extra \$5 for the same pair of jeans. In terms of consumer surplus loss, nine million consumers lose \$5 each for a total of \$45 million (area $a + b + c$), while the remaining one million lose a total of \$2.5 million (area d).

We can now ask whether a household would be willing to lobby the government to oppose the blue jeans tariff because of the extra cost they would incur. The likely answer is no. For most households, such a small price increase would hardly be noticed. Most consumers do not purchase blue jeans frequently. Also, blue jeans with different styles and brand names typically differ considerably in price. Consumers, who rarely keep track of events affecting particular markets, are unlikely to know that a tariff has even been implemented on the product considered or discussed.

If a person did know of an impending tariff, then presumably \$5 is the maximum a household would be willing to pay toward a lobbying effort, since that is the most one can gain if a tariff is prevented. One might argue that if even a fraction of the \$5 could be collected from some portion of the ten million consumer households, millions of dollars could be raised to contribute to an opposition lobbying effort. However, collecting small contributions from such a large group would be very difficult to do effectively.

Consider the problems one would face in spearheading a consumer lobbying effort to oppose the blue jeans tariff in this example. A seemingly reasonable plan would be to collect a small amount of money from each household hurt by the tariff and use those funds to pay for a professional lobbying campaign directed at the key decision makers. The first problem faced is how to identify which households are

likely to be affected by the tariff. Perhaps many of these households purchased blue jeans last year, but many others may be new to the market in the upcoming year. Finding the right people to solicit money from would be a difficult task.

Even if you could identify them, you would have to find a way to persuade them that they ought to contribute. Time spent talking to each household has an opportunity cost to the household member since that person could be doing something else. Suppose that a person values her time at the hourly wage rate that she earns at her job. If she makes \$20 per hour, then you'll have less than fifteen minutes to convince her to contribute to the lobbying effort since fifteen minutes is worth the \$5 you are trying to save for her. The point here is that even learning about the problem is costly for the household. For small savings, a lobbying group will have to convince its contributors very quickly.

Suppose we knew the names and addresses of the ten million affected households. Perhaps we could send a letter to each of them with a stamped return envelope asking to return it with a \$2 or \$3 contribution to the lobbying effort. With this plan, even purchasing the stamps to mail the envelopes would cost \$3,400,000. One would need to get over half of the households to send in \$3 each just to cover the cost of the mailing. Recipients of the letters will reasonably question the trustworthiness of the solicitation. Will the money really be put to good use? The chances of getting any more than a small return from this kind of solicitation is highly unlikely.

If contributions can be collected, the lobbying group will face another problem that arises with large groups: free ridership. **Free riding**⁶ occurs when someone enjoys the benefits of something without paying for it. The lobbying effort, if successful, will benefit all blue jeans consumers regardless of whether they contributed to the lobbying campaign. In economic terms, we say that the lobbying effort is a public good because individual households cannot be excluded from the benefits of successful lobbying. One of the key problems with public good provision is that individuals may be inclined to free ride—that is, to obtain the benefit without having contributed to its provision. Those who do not contribute also get the added benefit of the full \$5 surplus if the lobbying campaign is successful.

The main point of this discussion, though, is that despite the fact that \$47.5 million dollars will be lost to consumers of blue jeans if the \$5 tariff is implemented, it is very unlikely that this group will be able to form a lobbying campaign to oppose the tariff. Since each household will lose \$5 at most, it is extremely unlikely for any reasonable person to spend sufficient time to mount a successful lobbying campaign. Even if one person or group decided to spearhead the effort and collect contributions from others, the difficulties they would face would likely be

6. When someone enjoys the benefits of something without paying for it, especially when the product is a public good.

insurmountable. In the end, government decision makers would probably hear very little in the way of opposition to a proposed tariff.

Many of the arguments are discussed in detail in Mancur Olson's well-known book *The Logic of Collective Action*. One of the book's key points is that large groups are much less effective than small groups in applying effective lobbying pressure on legislators.

KEY TAKEAWAYS

- Although the loss of consumer surplus is the largest welfare effect of a tariff, because the number of consumers affected is also very large, the effect on each consumer is relatively small.
- Because a large number of consumers are affected to a small degree, it is difficult to identify precisely which consumers are affected by the tariff.
- Because the per-consumer cost of the tariff is low, most individual lobbying efforts to protest the tariff will cost the individual more than the cost of the tariff.
- Large groups are much less effective than small groups in applying effective lobbying pressure on legislators.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe when a person receives a benefit, especially a public good, without contributing a fair share to pay for it.
 - b. Of concentrated or dispersed, this is how to describe the typical losses that accrue to consumers because of an import tariff.
 - c. The amount of money lost by each consumer of coffee due to a \$0.25 tariff if \$35 million is lost in consumer surplus in a market of seventy million consumers.
 - d. Of small or large, this sized group is more likely to form an effective lobby.

10.5 The Producers' Lobbying Decision

LEARNING OBJECTIVE

1. Learn the lobbying implications of the concentrated benefits of protection to producer interests.

On the producers' side, let's assume that there are thirty-five separate, and equally sized, firms. If a \$5 tariff is implemented, producers as a group would gain \$35 million in producer surplus. That means each firm stands to gain \$1 million. Domestic producers would also supply two million additional pairs of jeans, and that would require expansion of the industry labor force. Clearly, the tariff would be beneficial to the firm owners and to industry workers. The potential to expand production, add workers, and increase profits by \$1 million per firm will provide a strong motivation to participate in a lobbying effort. In the case of the firms, however, organization of a lobbying effort will be much easier than the opposing effort by consumers.

First of all, the \$1 million surplus accruing to each firm is pure gravy. Payments to workers and other factors are not a part of the \$1 million additional surplus; thus it is money over and above the marginal costs of additional production. For this reason, profit received in this manner is often referred to as "economic rents." Since the rents are concentrated in a small number of firms, with \$1 million going to each, each firm will have a strong incentive to participate in a lobbying campaign. But who's going to spearhead the effort?

Organization of a lobbying campaign will probably be easier for firms than for consumers. First, the industry may have an industry association that maintains continual links with policymakers in state and federal governments. The workers in the industry might also belong to a trade union, which would also have interests in supporting a lobbying effort. Or a few of the industry leaders could take it upon themselves to begin the effort (although that is assumed away in the example). Second, as a smaller group, it is easy to identify the likely beneficiaries from the tariff and to solicit contributions. The lobbying group should easily be able to collect millions of dollars to support an extensive lobbying effort. A mere contribution of \$50,000 per firm would generate \$1.75 million that could be used to hire a professional lobbying team. Even if the chances of a successful outcome are small, it may still be practical for the firms to contribute to a lobbying effort. The return on that \$50,000 "investment" would be \$1 million if successful. That's a 2,000

percent rate of return—much higher than any brick-and-mortar investment project that might be considered. Free riding would also be less likely to occur since with only thirty-five firms to keep track of, contributors would probably learn who is not participating. Nonparticipation would establish a poor reputation for a firm and could have unpleasant consequences in its future industry association dealings.

With a well-financed lobbying effort, it would not be difficult to make decision makers aware that there is resounding support for the tariff within the industry community. Newspaper and television ads could be purchased to raise public awareness. Interested parties could be flown to the capitol to speak with key decision makers. In this way, the chances of obtaining the tariff may be increased substantially.

The Mancur Olson result applies in reverse to small groups. Small groups are much more effective than large groups in applying effective lobbying pressure on legislators.

KEY TAKEAWAYS

- Although the increase in producer surplus is a small welfare effect in the example, because the number of producers affected is relatively small, the effect of the tariff on each producer is relatively large.
- Because a small number of producers are affected to a sizeable degree, it is easy to identify who is positively affected by the tariff.
- Because the per-producer benefit of the tariff is high, firm lobbying efforts to promote the tariff will likely be a worthwhile investment.
- Small groups are much more effective than large groups in applying effective lobbying pressure on legislators.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *concentrated* or *dispersed*, this is how to describe the typical benefits that accrue to producers because of an import tariff.
 - b. The amount of money gained by each producer of coffee due to a \$0.25 tariff if \$20 million is gained in producer surplus in a market of twenty producing firms.
 - c. The rate of return on a \$50,000 lobbying expense if that lobbying results in a \$0.30 tariff on coffee and nets the firm an additional \$3,000,000 in profit.
 - d. Of *consumer groups* or *producer groups*, this group is more likely to form an effective lobby.

10.6 The Government's Decision

LEARNING OBJECTIVE

1. Learn how the lobbying and tariff revenue implications of a tariff affect the decision of the government.

How the government decides whether to offer the \$5 tariff, and who decides, will depend on the procedural rules of the democratic country in question. The tariff might be determined as a part of an administered procedure, such as a safeguards action or an antidumping action. Or the tariff may be determined as a part of a bill to be voted on by the legislature and approved, or not, by the executive. Rather than speaking about a particular type of government action, however, we shall consider the motivations of the government more generically.

The first thing the government may notice when being petitioned to consider raising the tariff is that government revenues will rise by \$5 million. Relative to many government budgets, this is a small amount, and so it may have very little influence on policymakers' decision. However, it will help reduce a budget deficit or add to the monies available for spending on government programs. Thus it could have a small influence.

In a democratic society, governments are called on to take actions that are in the interests of their constituents. If government officials, in this example, merely listen to their constituents, one thing should be obvious. The arguments of the industry seeking protection will surely resonate quite loudly, while the arguments of the consumers who should be opposed to the tariff will hardly even be heard. If a government official bases his or her decision solely on the "loudness" of the constituents' voices, then clearly he or she would vote for the tariff. This is despite the fact that the overall cost of the tariff to consumers outweighs the benefits to the industry and the government combined.

Notice that the decision to favor the tariff need not be based on anything underhanded or illegal on the part of the industry lobbyists. Bribes need not be given to secure votes. Nor does the industry lobby need to provide false or misleading information. Indeed, the lobby group could provide flawlessly accurate information and still win the support of the officials. Here's why.

It would be natural for the industry lobby group to emphasize a number of things. First, jobs would be saved (or created) as a result of the tariff. If a number can be attached, it will be. For example, suppose the industry supported 25,000 jobs in the initial equilibrium, when eight million pairs of jeans were produced by the domestic industry. That averages to 320 jeans produced per worker. Thus, when the industry cuts production by two million units, it amounts to 6,250 jobs. The lobby group could then frequently state the “fact” that the tariff will create 6,250 jobs. Second, the lobby would emphasize how the tariff would restore the vitality of the industry. If a surge of imports contributed to the problem, then the lobby would undoubtedly blame foreign firms for taking jobs away from hardworking domestic citizens. Finally, the lobby would emphasize the positive government budget effects as a result of the tariff revenue. All of this information clearly would be quite true.

If the lobby mentioned the higher prices that would result from the tariff, surely it would argue it is a small price to pay to save so many jobs. The lobby might even convince consumers of blue jeans that it is worth paying extra for jeans because it will save domestic jobs. After all, perhaps their own jobs will one day be in jeopardy due to imports. Plus, it is such a small price to pay: at only \$5 extra, no one will even notice!

For a politician facing potential reelection, there is another reason to support the industry over the consumers, even with full information about the effects. Support of the industry will probably generate more future votes. Here’s why.

First, since industry members—management and workers—have a bigger stake in the outcome, they will be more likely to remember the politician’s support (or lack of support) on this issue at election time. Second, the politician can use his support for the industry in his political ads. Consider this political ad if he supports the industry: “I passed legislation that created over six thousand jobs!” Compare it with this truthful ad if he doesn’t support the industry: “By opposing protectionist legislation, I saved you five bucks on blue jeans!” Which one do you think sounds better?

KEY TAKEAWAY

- If representatives in a democracy base policy choices on the interests of their constituents and if industry lobbyists are more organized and “vocal” in their demands than consumers, then governments will more likely choose policies like tariffs.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *consumer voices* or *producer voices*, these are more likely to be heard by government officials in a representative democracy.
 - b. Of *consumer interests* or *producer interests*, governments in representative democracies are more likely to implement policies favoring these.

10.7 The Lobbying Problem in a Democracy

LEARNING OBJECTIVE

1. Recognize some of the problems and pitfalls of policy choice in a representative democracy.

There is a real problem with the lobbying process in democratic societies. Even though lobbying is a legitimate process of information transfer between constituents and government decision makers, it also produces some obvious disparities. Whenever policy actions generate concentrated benefits and dispersed costs, the incentives and abilities to lobby are significantly different across groups. Potential beneficiaries can often use the advantage of small group size and large potential windfalls to wield disproportionate influence on decision makers. Potential losers, whose numbers are large and whose expected costs per person are quite small, have almost no ability to lobby the government effectively. Thus, in a democratic society in which lobbying can influence decisions, decisions are likely to be biased in the favor of those policies that generate concentrated benefits and dispersed losses.

Unfortunately, and perhaps coincidentally, most policy actions taken produce concentrated benefits and dispersed losses. In the case of trade policies, most protectionist actions will cause concentrated benefits to accrue to firms, whereas losses will be dispersed among millions of consumers. This means that protectionist policies are more likely to win political support, especially when lobbying can directly affect legislated actions. Protectionism can easily occur even though the sum total effects of the policy may be negative.

In many countries, a protectionist tendency is reflected in the type of trade policy procedures that are available by law. Escape clause, antisubsidy, and antidumping policies are examples of laws designed to protect firms and industries in particular situations. In evaluating these types of petitions in the United States, there is no requirement that effects on consumers be considered in reaching a decision. Clearly, these laws are designed to protect the concentrated interests of producing firms. It would not be surprising, and indeed it seems likely, that the concentrated interests of businesses affected the ways in which the laws were originally written. The absence of a consumer lobby would also explain why consumer effects are never considered in these actions.

KEY TAKEAWAY

- Democratic governments are more likely to choose policies that generate concentrated benefits and dispersed losses, regardless of whether the sum total effects are positive or negative.

EXERCISES

1. Suppose a small country implements a tariff on chicken imports. In the table below indicate whether each group is a winner or loser and whether the effects on that group are concentrated or dispersed.

TABLE 10.2 POLITICAL ECONOMY EFFECTS OF A TARIFF

Name of Group	Winners or Losers	Concentrated or Dispersed
Chicken Producers		
Chicken Consumers		
Taxpayers or Recipients of Government Benefits		

2. Suppose a small country implements an export subsidy on soybeans. In the table below indicate whether each group is a winner or loser and whether the effects on that group are concentrated or dispersed.

TABLE 10.3 POLITICAL ECONOMY EFFECTS OF AN EXPORT SUBSIDY

Name of Group	Winners or Losers	Concentrated or Dispersed
Soybean Producers		
Soybean Consumers		
Taxpayers or Recipients of Government Benefits		

Chapter 11

Evaluating the Controversy between Free Trade and Protectionism

Perhaps the most important policy issue of an international trade course is to answer the question “Should a country pursue free trade or some type of selected protection?” Academics, philosophers, policy analysts, and legislators have addressed this question for hundreds of years. And unfortunately, there is still no definitive answer.

The reason is that both free trade and selected protection have both positive and negative aspects. No one policy choice is clearly superior. Nonetheless, economists who have studied trade theory and policy tend to support free trade more so than just about any other contentious economic policy under public consideration. The reasons for this near consensus are complex and poorly understood by the general public. This chapter explains the economic case for free trade through the lens of trade theory and argues that even though free trade may not be “optimal,” it is nonetheless the most pragmatic policy option a country can follow.

11.1 Introduction

LEARNING OBJECTIVE

1. Understand the basis for the modern support for free trade among economists.

For hundreds of years, at least since Adam Smith's publication of *The Wealth of Nations*, the majority of economists have been strong supporters of free trade among nations. Paul Krugman once wrote that if there were an economist's creed, it would surely contain the affirmation, "I advocate free trade." See Paul Krugman, "Is Free Trade Passe?" *Journal of Economic Perspectives* 1, no. 2 (1987): 131–44.

The original arguments for free trade began to supplant mercantilist views in the early to mid-eighteenth century. Many of these original ideas were based on simple exchange or production models that suggested that free trade would be in everyone's best interests and surely in the national interest. During the nineteenth and twentieth centuries, however, a series of objections were raised suggesting that free trade was not in everyone's interest and perhaps was not even in the national interest. The most prominent of these arguments included the infant industry argument, the terms of trade argument, arguments concerning income redistribution, and more recently, strategic trade policy arguments. Although each of these arguments might be thought of as weakening the case for free trade, instead, each argument brought forth a series of counterarguments that have acted to reassert the position of free trade as a favored policy despite these objections. The most important of these counterarguments include the potential for retaliation, the theory of the second best, the likelihood of incomplete or imperfect information, and the presence of lobbying in a democratic system.

What remains today is a modern, sophisticated argument in support of free trade among nations. It is an argument that recognizes that there are numerous exceptions to the notion that free trade is in everyone's best interests. The modern case for free trade does not contend, however, that these exceptions are invalid or illogical. Rather, it argues that each exception supporting government intervention in the form of a trade policy brings with it additional implementation problems that are likely to make the policy impractical.

Before presenting the modern argument, however, it is worth deflecting some of the criticisms that are sometimes leveled against the economic theory of free trade.

For example, the modern argument for free trade is not based on a simplistic view that everyone benefits from free trade. Indeed, trade theory, and experience in the real world, teaches us that free trade, or trade liberalization, is likely to generate losers as well as winners.

The modern argument for free trade is not based on unrealistic assumptions that lead to unrealistic conclusions. Although it is true that many assumptions contained within any given trade model do not accurately reflect many realistic features of the world, the modern argument for free trade is not based on the results from any *one* model. Instead, the argument is based on a collection of results from numerous trade models, which are interpreted in reference to realistic situations. If one considers the collection of all trade models jointly, it is much more difficult to contend that they miss realistic features of the world. Trade theory (as a collection of models) does consider imperfectly competitive markets, dynamic effects of trade, externalities in production and consumption, imperfect information, joint production, and many other realistic features. Although many of these features are absent in any one model, they are not absent from the joint collection of models, and it is this “extended model” that establishes the argument for free trade. Ideally, we would create a supermodel of the world economy that simultaneously incorporates all realistic features of the world and avoids what are often called “simplifying assumptions.” Unfortunately, this is not a realistic possibility. As anyone who has studied models of the economy knows, even models that are very simple in structure can be extremely difficult to comprehend, much less solve. As a result, we are forced to “interpret” the results of simple models as we apply them to the complex real world.

KEY TAKEAWAY

- The modern support for free trade by most economists is based on a collection of results from a collection of models that incorporate many realistic features of the world into the analysis.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The statement suggested by Paul Krugman as being an element of the economist’s creed—if ever there were such a thing.
 - b. This is who will benefit from free trade according to a simplistic view held by some free trade advocates.
 - c. This is what causes unrealistic conclusions in trade theory according to some free trade opponents.
 - d. The conclusions of *one* model of international trade or *many* models of international trade are best used to make trade policy prescriptions.

11.2 Economic Efficiency Effects of Free Trade

LEARNING OBJECTIVE

1. Learn the major source of support for free trade across a variety of trade models.

The main source of support for free trade lies in the positive production and consumption efficiency effects. In every model of trade, there is an improvement in aggregate production and consumption efficiency when an economy moves from autarky to free trade. This is equivalent to saying that there is an increase in national welfare. This result was demonstrated in the Ricardian model, the immobile factor model, the specific factor model, the Heckscher-Ohlin model, the simple economies-of-scale model, and the monopolistic competition model. The result can also be shown if there are differences in demand between countries. Each of these models shows that a country is likely to have greater national output and superior choices available in consumption as a result of free trade.

Production Efficiency

Improvements in production efficiency mean that countries can produce more goods and services with the same amount of resources. In other words, productivity increases for the given resource endowments available for use in production.

In order to achieve production efficiency improvements, resources must be shifted between industries within the economy. This means that some industries must expand while others contract. Exactly which industries expand and contract will depend on the underlying stimulus or basis for trade. Different trade models emphasize different stimuli for trade. For example, the Ricardian model emphasizes technological differences between countries as the basis for trade, the factor proportions model emphasizes differences in endowments, and so on. In the real world, it is likely that each of these stimuli plays some role in inducing the trade patterns that are observed.

Thus as trade opens, either the country specializes in the products in which it has a comparative technological advantage, or production is shifted to industries that use the country's relatively abundant factors most intensively, or production is shifted to products in which the country has relatively less demand compared with the rest

of the world, or production shifts to products that exhibit economies of scale in production.

If production shifts occur for any of these reasons, or for some combination of these reasons, then trade models suggest that total production would rise. This would be reflected empirically in an increase in the country's gross domestic product (GDP). This means that free trade would cause an increase in the level of the country's national output and income.

Consumption Efficiency

Consumption efficiency improvements arise for an individual when changes in the relative prices of goods and services allow the consumer to achieve a higher level of utility. Since the change in prices alters the choices a consumer has, we can say that consumption efficiency improvements imply that more satisfying choices become available. When multiple varieties of goods are available in a product category, as in the monopolistic competition model, then consumption efficiency improvements can mean that the consumer is able to consume greater varieties or is able to purchase a variety that is closer to his ideal.

Although improvements in consumption efficiency are easy to describe for an individual consumer, it is much more difficult to describe consumption efficiency conceptually for the aggregate economy. Nevertheless, when aggregate indifference curves are used to describe the gains from trade, it is possible to portray an aggregate consumption efficiency improvement. One must be careful to interpret this properly, though. The use of an aggregate indifference curve requires the assumptions that (1) all consumers have identical preferences and (2) there is no redistribution of income as a result of the changes in the economy. We have seen, however, that in most trade models income redistribution will occur as an economy moves to free trade, and it may be impossible to redistribute afterward. It is also likely that individuals have different preferences for goods, which also weakens the results using aggregate indifference curves.

KEY TAKEAWAYS

- The main sources of support for free trade are the positive production and consumption efficiency effects that arise in numerous models when countries trade freely.
- Production efficiency improvements mean that countries produce more goods and services with the same amount of resources.
- Consumption efficiency improvements mean that countries consume a more satisfying mix of goods and services.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term often used as a synonym for an improvement in economic efficiency.
 - b. The type of efficiency improvement in which productivity rises for the given resource endowments available for use in production.
 - c. The type of efficiency improvement relating to consumer choice adjustments in response to a policy change.
 - d. The enhancement of *this* is what many economic models show will arise by moving to free trade.

11.3 Free Trade and the Distribution of Income

LEARNING OBJECTIVES

1. Recognize that a movement to free trade will cause a redistribution of income within the country.
2. Understand how compensation can relieve the problems caused by income redistribution.

A valid criticism of the case for free trade involves the issue of income distribution. Although most trade models suggest that aggregate economic efficiency is raised with free trade, these same models do not indicate that every individual in the economy will share in the benefits. Indeed, most trade models demonstrate that movements to free trade will cause a redistribution of income between individuals within the economy. In other words, some individuals will gain from free trade while others will lose. This was seen in the immobile factor model, the specific factor model, the Heckscher-Ohlin model, and the partial equilibrium analysis of trade liberalization.

There have been two general responses by economists concerning the income distribution issue. Some have argued that the objective of economics is solely to determine the most efficient policy choices. Introductory textbooks often suggest that the objective of the economics discipline is to determine how to allocate scarce resources toward production and consumption. Economists describe an allocation as “optimal” when it achieves the maximum level of aggregate economic efficiency. Put in these terms, economic analysis is “positive” in nature. Positive economics refers to studies that seek to answer questions pertaining to how things work in the economy and the subsequent effects. Positive economic analysis does not intend to explain what “should” be done. Issues pertaining to income distribution are commonly thought of as “normative” in nature, in that the concern is often over what the distribution “should” be. If we apply this reasoning to international trade, then, issues such as the appropriate income distribution are beyond the boundaries of the discipline and should be left to policymakers, government officials, or perhaps philosophers to determine.

Perhaps a more common response by economists concerning the income distribution issue is to invoke the compensation principle. A substantial amount of work by economists has been done to show that because free trade causes an increase in economic efficiency, it is generally possible to redistribute income from

the winners to the losers such that, in the end, every individual gains from trade. The basic reason this is possible is that because of the improvement in aggregate efficiency, the sum of the gains to the winners exceeds the sum of the losses to the losers. This implies that it is theoretically possible for the potential winners from free trade to bribe the losers and leave everyone better off as a result of free trade. This allows economists to argue that free trade, coupled with an appropriate compensation package, is preferable to some degree of protectionism.

One major practical problem with compensation, however, is the difficulty of implementing a workable compensation package. In order to achieve complete compensation, one must be able to identify not only who the likely winners and losers will be but also how much they will win and lose and when in time the gains and losses will accrue. Although this is relatively simple to do in the context of a single trade model, such as the Heckscher-Ohlin model, it would be virtually impossible to do in practice given the complexity of the real world. The real world consists of tens of thousands of different industries producing millions of products using thousands of different factors of production. The sources of trade are manifold, including differences in technology, endowments, and demands, as well as the presence of economies of scale. Each source of trade, in turn, stimulates a different pattern of income redistribution when trade liberalization occurs. In addition, the pattern of redistribution over time is likely to be affected by the degree of mobility of factors between industries as the adjustment to free trade occurs. This was seen in the context of simple trade models, from the immobile factor model to the specific factor model to the Heckscher-Ohlin model.

Even in the context of simple trade models, a workable compensation mechanism is difficult to specify. An obvious solution would seem to be for the government to use taxes and subsidies to facilitate compensation. For example, the government could place taxes on those who would gain from free trade (or trade liberalization) and provide subsidies to those who would lose. However, if this were implemented in the context of many trade models, then the taxes and subsidies would change the production and consumption choices made in the economy and would act to reduce or eliminate the efficiency gains from free trade. The government taxes and subsidies, in this case, represent a policy-imposed distortion that, by itself, reduces aggregate economic efficiency. If the compensation package reduces efficiency more than the movement to free trade enhances efficiency, then it is possible for the nation to be worse off in free trade when combined with a tax/subsidy redistribution scheme. Dixit and Norman (1980) showed that under some conditions it is possible to specify a tax and subsidy policy that would guarantee an increase in aggregate economic efficiency with free trade. See A. Dixit and V. Norman, *Theory of International Trade: A Dual General Equilibrium Approach* (Cambridge: Cambridge University Press, 1980). The simple way to eliminate this problem, conceptually, is to suggest that the redistribution take place as a “lump-sum” redistribution. A

lump-sum redistribution¹ is one that takes place after the free trade equilibrium is reached—that is, after all production and consumption decisions are made but before the actual consumption takes place. Then, as if in the middle of the night when all are asleep, goods are taken away from those who have gained from free trade and left at the doors of those who had lost. Lump-sum redistributions are analogous to Robin Hood stealing from the rich and giving to the poor. As long as this redistribution takes place after the consumption choices have been made and without anyone expecting a redistribution to occur, then the aggregate efficiency improvements from free trade are still realized. Of course, although lump-sum redistributions are a clever conceptual or theoretical way to “have your cake and eat it too,” it is not practical or workable in the real world.

This implies that although compensation can solve the problem of income redistribution at the theoretical level, it is unlikely that it will ever solve the problem in the real world. Although some of the major gains and losses from free trade may be identifiable and quantifiable, it is unlikely that analysts would ever be able to identify all who would gain and lose in order to provide compensation and assure that everyone benefits. This means that free trade is extremely likely to cause uncompensated losses to some individuals in the economy. To the extent that these individuals expect these losses and can measure their expected value (accurately or not), then there will also likely be continued resistance to free trade and trade liberalization. This resistance is perfectly valid. After all, trade liberalization involves a government action that will cause injury to some individuals for which they do not expect to be adequately compensated. Furthermore, the economic efficiency argument will not go very far to appease these groups. Would you accept the argument that your expected losses are justifiable because others will gain more than you lose?

One final argument concerning the compensation issue is that compensation to the losers may not even be justifiable. This argument begins by noting that those who would lose from free trade are the same groups who had gained from protectionism. Past protectionist actions represent the implementation of government policies that had generated benefits to certain selected groups in the economy. When trade liberalization occurs, then, rather than suggesting that some individuals lose, perhaps it is more accurate to argue that the special benefits are being eliminated for those groups. On the other hand, those groups that benefit from free trade are the same ones that had suffered losses under the previous regime of protectionism. Thus their gains from trade can be interpreted as the elimination of previous losses. Furthermore, since the previous protectionist actions were likely to have been long lasting, one could even argue that the losers from protection (who would gain from free trade) deserve to be compensated for the sum total of their past losses. This would imply that upon moving to free trade, a redistribution ought to be made not from the winners in trade to the losers but

1. A redistribution of income that takes place after the free trade equilibrium is reached—that is, after all production and consumption decisions are made but before the actual consumption takes place.

from the losers in trade to the winners. Only in this way could one make up for the transgressions of the past. As before, though, identifying who lost and who gained and by how much would be virtually impossible to achieve, thus making this compensation scheme equally unworkable.

KEY TAKEAWAYS

- One major problem with movements to free trade is the redistribution of income described in many trade models. This means that although some individuals will benefit from free trade, many others will lose.
- One way to deflect the redistribution concern is to argue that economic analysis provides the positive results of trade policies and is not intended to answer the normative questions of what should be done.
- Another way to deflect the concern about income redistribution is to support compensation from the winners to the losers to assure that all parties benefit from free trade.
- Because compensation requires an enormous amount of information about who wins and loses from trade, how much they win and lose, and when they win and lose, it is impractical to impossible to completely compensate the losers from free trade in a real-world setting.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. A principle that, if applied in practice, could eliminate the negative impacts of income redistribution that may arise with free trade.
 - b. This is what many trade models show will happen to national income because of trade liberalization.
 - c. This type of compensation can avoid affecting consumption and production decisions.
 - d. The compensation using these two government policies is likely to affect production and consumption decisions.
 - e. The name of the mythical character best associated with lump-sum compensation.
 - f. Of *a little* or *a lot*, this is how much information the government needs to make compensation effective.

11.4 The Case for Selected Protection

LEARNING OBJECTIVE

1. Identify the cases in which the implementation of selected protectionism, targeted at particular industries with particular goals in mind, could raise national welfare.

An argument for **selected protection**² arises in the presence of imperfectly competitive markets, market distortions, or both. In these cases, it is often possible to show that an appropriately targeted trade policy (selected protection) can raise aggregate economic efficiency. In other words, free trade need not always be the best policy choice when the objective is to maximize national welfare. Numerous examples found in the trade literature demonstrate that selected protectionism applied under certain circumstances can raise national welfare. These results are in contrast with the standard trade models, which show that free trade is the best policy to maximize economic efficiency. The reason for the conflict is that the standard trade models, in most cases, explicitly assume that markets are perfectly competitive and implicitly assume there are no market distortions.

This general criticism of the standard case for free trade begins by noting that the real world is replete with examples of market imperfections and distortions. These include the presence of externalities both static and dynamic, both positive and negative, and in both production and consumption; markets in which production takes place with monopolistic or oligopolistic firms making positive profits; markets that do not clear, as when unemployment arises; the presence of public goods; the presence of imperfect or asymmetric information; the presence of distorting government policies and regulations; and the presence of national market power in international markets. When these features are included in trade models, it is relatively easy to identify trade policies that can sufficiently correct the market imperfection or distortion so as to raise aggregate efficiency.

2. A trade policy that is appropriately selected so as to raise national welfare in a market containing market imperfections or distortions.

For example, an optimal tariff or optimal quota set by a country that is large in an international import market can allow the nation to take advantage of its monopsony power in trade and cause an increase in national welfare. Similarly, an optimal export tax or voluntary export restraint (VER) set by a large country in an international export market will allow it to take advantage of its monopoly power in trade and generate an increase in welfare. This argument for protection is known as the “terms of trade argument.”

A tariff applied to protect an import-competing industry from a surge in foreign imports may reduce or eliminate the impending unemployment in the industry. If the cost of unemployment to the affected workers is greater than the standard net national welfare effect of the tariff, then the tariff may improve national welfare.

A tariff used to restrict imports of goods from more-efficient foreign firms may sufficiently stimulate learning effects within an industry to cause an increase in productivity that, in time, may allow the domestic firms to compete with foreign firms—even without continued protection. These learning effects—in organizational methods, in management techniques, in cost-cutting procedures—might in turn spill over to other sectors in the economy, stimulating efficiency improvements in many other industries. All together, the infant industry protection may cause a substantial increase in the growth of the gross domestic product (GDP) relative to what might have occurred otherwise and thus act to improve national welfare.

A tariff used to stimulate domestic production of a high-technology good might spill over to the research and development division and cause more timely innovations in next-generation products. If these firms turn into industry leaders in these next-generation products, then they will enjoy the near-monopoly profits that accrue to the original innovators. As long as these long-term profits outweigh the short-term costs of protection, national welfare may rise.

An import tariff applied against a foreign monopoly supplying the domestic market can effectively shift profits from the foreign firm to the domestic government. Despite the resulting increase in the domestic price, national welfare may still rise. Also, export subsidies provided to domestic firms that are competing with foreign firms in an oligopoly market may raise domestic firms' profits by more than the cost of the subsidy, especially if profits can be shifted away from the foreign firms. These two cases are examples of a strategic trade policy.

If pollution, a negative production externality, caused by a domestic import-competing industry is less than the pollution caused by firms in the rest of the world, then a tariff that restricts imports may sufficiently raise production by the domestic firm relative to foreign firms and cause a reduction in world pollution. If the benefits that accrue due to reduced worldwide pollution are greater than the standard cost of protection, then the tariff will raise world welfare.

Alternatively, if pollution is caused by a domestic export industry, then an export tax would reduce domestic production along with the domestic pollution that the production causes. Although the export tax may act to raise production and pollution in the rest of the world, as long as the domestic benefits from the

pollution reduction outweigh the costs of the export tax, domestic national welfare may rise.

If certain domestically produced high-technology goods could wind up in the hands of countries that are our potential enemies, and if these goods would allow those countries to use the products in a way that undermines our national security, then the government could be justified to impose an export prohibition on those goods to those countries. In this case, if free trade were allowed in these products, it could reduce the provision of a public good, namely, national security. As long as the improvement in national security outweighs the cost of the export prohibition, national welfare would rise.

These are just some of the examples (many more are conceivable) in which the implementation of selected protectionism, targeted at particular industries with particular goals in mind, could act to raise national welfare, or aggregate economic efficiency. Each of these arguments is perfectly valid conceptually. Each case arises because of an assumption that some type of market imperfection or market distortion is present in the economy. In each case, national welfare is enhanced because the trade policy reduces or eliminates the negative effects caused by the presence of the imperfection or distortion and because the reduction in these effects can outweigh the standard efficiency losses caused by the trade policy.

It would seem from these examples that a compelling case can certainly be made in support of selected protectionism. Indeed, Paul Krugman (1987) wrote that “the case for free trade is currently more in doubt than at any time since the 1817 publication of [David] Ricardo’s *Principles of Political Economy*.¹ See Paul Krugman, “Is Free Trade Passe?” *Journal of Economic Perspectives* 1, no. 2 (1987): 131–44. Many of the arguments showing the potential for welfare-improving trade policies described above have been known for more than a century. The infant industry argument can be traced in the literature as far back as a century before Adam Smith argued against it in *The Wealth of Nations* (1776). The argument was later supported by writers such as Friedrich List in *The National System of Political Economy* (1841). See Friedrich List, *The National System of Political Economy*, McMaster University Archive for the History of Economic Thought, <http://socsci2.socsci.mcmaster.ca:80/~econ/ugcm/3ll3/list/index.html>. and John Stuart Mill in his *Principles of Political Economy* (1848). See John Stuart Mill, *Principles of Political Economy*, McMaster University Archive for the History of Economic Thought, <http://socsci2.socsci.mcmaster.ca:80/~econ/ugcm/3ll3/mill/index.html>. The terms of trade argument was established by Robert Torrens in 1844 in *The Budget: On Commercial and Colonial Policy*. See Robert Torrens, *The Budget: On Commercial and Colonial Policy* (London: Smith, Elder, 1844). Frank Graham, in his 1923 article “Some Aspects of Protection Further Considered,” noted the possibility that free trade would reduce welfare if there were variable returns to scale in

production. See Frank Graham, “Some Aspects of Protection Further Considered,” *The Quarterly Journal of Economics* 37, no. 2 (February 1923): 199–227. During the 1950s and 1960s, market distortions such as factor-market imperfections and externality effects were introduced and studied in the context of trade models. The strategic trade policy arguments are some of the more recent formalizations showing how market imperfections can lead to welfare-improving trade policies. Despite this long history, economists have generally continued to believe that free trade is the best policy choice. The main reason for this almost unswerving support for free trade is because as arguments supporting selected protectionism were developed, equally if not more compelling counterarguments were also developed.

KEY TAKEAWAYS

- In the presence of market imperfections or distortions, selected protection can often raise a country’s national welfare.
- Because real-world markets are replete with market imperfections and distortions, free trade is not the optimal policy to improve national welfare.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe market conditions that open up the possibility for welfare-improving trade policies.
 - b. The term used to describe a market equilibrium in which market imperfections or distortions are present.
 - c. Of *very many* or *very few*, this is the amount of market imperfections likely to be present in modern national economies.
 - d. Of *true* or *false*, a tariff can raise a nation’s welfare when it is a large importing country.
 - e. Of *true* or *false*, a tariff can raise national welfare in the presence of an infant industry.
 - f. Of *true* or *false*, a tariff can raise national welfare if all markets are perfectly competitive and if there are no market imperfections or distortions.
2. Identify a trade policy that can potentially raise national welfare in each of the following situations.
 - a. When a foreign monopoly supplies the domestic market with no import-competing producers.
 - b. When a domestic negative production externality is caused by a domestic industry that exports a portion of its production to the rest of the world.
 - c. When a positive production externality is caused by a domestic industry that competes with imports.
 - d. When a domestic negative consumption externality is caused by domestic consumers in a market in which the country exports a portion of its production to the rest of the world.
 - e. When a country is large in an export market.

11.5 The Economic Case against Selected Protection

LEARNING OBJECTIVE

1. Learn the valid counterarguments to the use of selected protection when market imperfections or distortions are present.

The economic case against selected protectionism does not argue that the reasons for protection are conceptually or theoretically invalid. Indeed, there is general acceptance among economists that free trade is probably not the best policy in terms of maximizing economic efficiency in the real world. Instead, the counterarguments to selected protectionism are based on four broad themes: (1) potential reactions by others in response to one country's protection, (2) the likely presence of superior policies to raise economic efficiency relative to a trade policy, (3) information deficiencies that can inhibit the implementation of appropriate policies, and (4) problems associated with lobbying within democratic political systems. We shall consider each of these issues in turn.

The Potential for Retaliation

One of the problems with using some types of selected protection arises because of the possibility of retaliation by other countries using similar policies. For example, it was shown that whenever a large country in the international market applies a policy that restricts exports or imports (optimally), its national welfare will rise. This is the terms of trade argument supporting protection. However, it was also shown that the use of an optimal trade policy in this context always reduces national welfare for the country's trade partners. Thus the use of an optimal tariff, export tax, import quota, or voluntary export restraint (VER) is a "beggar-thy-neighbor" policy—one country benefits only by harming others. For this reason, it seems reasonable, if not likely, that the countries negatively affected by the use of such policies, if they are also large in international markets, would retaliate by setting optimal trade policies restricting their exports and imports to the rest of the world. In this way, the retaliating country could generate benefits for itself in some markets to compensate for its losses in others.

However, the final outcome after retaliation occurs is very likely to be a reduction in national welfare for both countries. Harry Johnson (1953) showed the possibility that one country might still improve its national welfare even after a trade war (i.e., optimal protection followed by optimal retaliation); however, this seems an

unlikely outcome in real-world cases. Besides, even if one country did gain, it would still do so at the expense of its trade partners, which remains an unsavory result. See Harry G. Johnson, “Optimum Tariffs and Retaliation,” *Review of Economic Studies* 21, no. 2 (1953): 142–53. This occurs because each trade policy action results in a decline in world economic efficiency. The aggregate losses that accrue to one country as a result of the other’s trade policy will always exceed the benefits that accrue to the policy-setting country. When every large country sets optimal trade policies to improve its terms of trade, the subsequent reduction in world efficiency dominates any benefits that accrue due to its unilateral actions.

What this implies is that although a trade policy can be used to improve a nation’s terms of trade and raise national welfare, it is unlikely to raise welfare if other large countries retaliate and pursue the same policies. Furthermore, retaliation seems a likely response because maintenance of a free trade policy in light of your trade partner’s protection would only result in national aggregate efficiency losses. Indeed, Robert Torrens, the originator of the terms of trade argument, was convinced that a large country should maintain protective barriers to trade when its trade partners maintained similar policies. The case for unilateral free trade even when one’s trade partners use protective tariffs is only valid when a country is small in international markets.

Perhaps the best empirical support for this result is the experience of the world during the Great Depression of the 1930s. After the United States imposed the Smoot-Hawley Tariff Act of 1930, raising its tariffs to an average of 60 percent, approximately sixty countries retaliated with similar increases in their own tariff barriers. As a result, world trade in the 1930s fell to one-quarter of the level attained in the 1920s. Most economists agree that these tariff walls contributed to the length and severity of the economic depression. That experience also stimulated the design of the reciprocal trade liberalization efforts embodied in the General Agreement on Tariffs and Trade (GATT).

The issue of retaliation also arises in the context of strategic trade policies. In these cases, a trade policy can be used to shift profits from foreign firms to the domestic economy and raise domestic national welfare. The policies work in the presence of monopolistic or oligopolistic markets by raising the international market share for one’s own firms. The benefits to the policy-setting country arise only by reducing the profits of foreign firms and subsequently reducing those countries’ national welfare. One exception arises in the model by J. Eaton and G. Grossman, “Optimal Trade and Industrial Policy under Oligopoly,” *Quarterly Journal of Economics* 101, no. 2 (1986): 383–406. Thus one country’s gains are other country’s losses, and strategic trade policies can rightfully be called beggar-thy-neighbor policies. Since foreign firms would lose from our country’s policies, as before, it is reasonable to expect retaliation by the foreign governments. However, because these policies essentially

just reallocate resources among profit-making firms internationally, it is unlikely for a strategic trade policy to cause an improvement in world economic efficiency. This implies that if the foreign country did indeed retaliate, the likely result would be reductions in national welfare for both countries.

Retaliations would only result in losses for both countries when the original trade policy does not raise world economic efficiency. However, some of the justifications for protection that arise in the presence of market imperfections or distortions may actually raise world economic efficiency because the policy acts to eliminate some of the inefficiencies caused by the distortions. In these cases, retaliation would not pose the same problems. There are other problems, though.

The Theory of the Second Best

One of the more compelling counterarguments to potentially welfare-improving trade policies relies on the theory of the second best. This theory shows that when private markets have market imperfections or distortions present, it is possible to add another (carefully designed) distortion, such as a trade policy, and improve economic efficiency both domestically and worldwide. The reason for this outcome is that the second distortion can correct the inefficiencies of the first distortion by more than the inefficiencies caused by the imposed policy. In economist's jargon, the original distorted economy is at a second-best equilibrium. In this case, the optimal trade policy derived for an undistorted economy (most likely free trade) no longer remains optimal. In other words, policies that would reduce national welfare in the absence of distortions can now improve welfare when there are other distortions present.

This argument, then, begins by accepting that trade policies (protection) can be welfare improving. The problem with using trade policies, however, is that in most instances they are a second-best policy choice. In other words, there will likely be another policy—a domestic policy—that could improve national welfare at a lower cost than any trade policy. The domestic policy that dominates would be called a first-best policy. The general rule used to identify first-best policies is to use that policy that “most directly” attacks the market imperfection or distortion. It turns out that these are generally domestic production, consumption, or factor taxes or subsidies rather than trade policies. The only exceptions occur when a country is large in international markets or when trade goods affect the provision of a public good such as national security.

Thus the counterargument to selected protection based on the theory of the second best is that first-best rather than second-best policies should be chosen to correct market imperfections or distortions.

Since trade policies are generally second best while purely domestic policies are generally first best, governments should not use trade policies to correct market imperfections or distortions. Note that this argument does not contend that distortions or imperfections do not exist, nor does it assume that trade policies could not improve economic efficiency in their presence. Instead, the argument contends that governments should use the most efficient (least costly) method to reduce inefficiencies caused by the distortions or imperfections, and this is unlikely to be a trade policy.

Note that this counterargument to protection is also effective when the issue is income distribution. Recall that one reason countries may use trade policies is to achieve a more satisfying income distribution (or to avoid an unsatisfactory distribution). However, it is unlikely that trade policies would be the most effective method to eliminate the problem of an unsatisfactory income distribution. Instead, there will likely be a purely domestic policy that could improve income distribution more efficiently.

In the cases where a trade policy is first best, as when a country is large in international markets, this argument does not act as a counterargument to protection. However, retaliation remains a valid counterargument in many of these instances.

Information Deficiencies

The next counterargument against selected protectionism concerns the likely informational constraints faced by governments. In order to effectively provide infant industry protection, or to eliminate negative externality effects, stimulate positive externality effects, or shift foreign profits to the domestic economy, the government would need substantial information about the firms in the market, their likely cost structures, supply and demand elasticities indicating the effects on supply and demand as a result of price changes, the likely response by foreign governments, and much more. Bear in mind that although it was shown that selected protection *could* generate an increase in national welfare, it does not follow that any protection would *necessarily* improve national welfare. The information requirements arise at each stage of the government's decision-making process.

First, the government would need to identify which industries possess the appropriate characteristics. For example, in the case of infant industries, the government would need to identify which industries possess the positive learning externalities needed to make the protection work. Presumably, some industries would generate these effects, while others would not. In the case of potential unemployment in a market, the government would need to identify in which

industries facing a surge of imports the factor immobility was relatively high. In the case of a strategic trade policy, the government would have to identify which industries are oligopolistic and exhibit the potential to shift foreign profits toward the domestic economy.

Second, the government would need to determine the appropriate trade policy to use in each situation and set the tariff or subsidy at the appropriate level. Although this is fairly straightforward in a simple theoretical model, it may be virtually impossible to do correctly in a real-world situation. Consider the case of an infant industry. If the government identified an industry with dynamic intertemporal learning effects, it would then need to measure how the level of production would influence the size of the learning effects in all periods in the future. It would also need to know how various tariff levels would affect the level of domestic production. To answer this requires information about domestic and foreign supply and demand elasticities. Of course, estimates of past elasticities may not work well, especially if technological advances or preference changes occur in the future. All of this information is needed to determine the appropriate level of protection to grant as well as a timetable for tariff reduction. If the tariff is set too low or for too short a time, the firms might not be sufficiently protected to induce adequate production levels and stimulate the required learning effects. If the tariff is set too high or for too long a period, then the firms might become lazy. Efficiency improvements might not be made and the learning effects might be slow in coming. In this case, the production and consumption efficiency losses from the tariff could outweigh the benefits accruing due to learning.

This same information deficiency problem arises in every example of selected protection. Of course, the government would not need pinpoint accuracy to assure a positive welfare outcome. As demonstrated in the case of optimal tariffs, there would be a range of tariff levels that would raise national welfare above the level attained in free trade. A similar range of welfare-improving protection levels would also hold in all the other cases of selected protection.

However, there is one other informational constraint that is even ignored in most economic analyses of trade policies. This problem arises when there are multiple distortions or imperfections present in the economy simultaneously (exactly what we would expect to see in the real world). Most trade policy analyses incorporate one economic distortion into a model and then analyze what the optimal trade policy would be in that context. Implicitly, this assumes either that there are no other distortions in the economy or that the market in which the trade policy is being considered is too small to have any external effects on other markets. The first assumption is clearly not satisfied in the world, while the second is probably not valid for many large industries.

The following example suggests the nature of the informational problem. Suppose there are two industries that are linked together because their products are substitutable in consumption to some degree. Suppose one of these industries exhibits a positive dynamic learning externality and is having difficulty competing with foreign imports (i.e., it is an infant industry). Assume the other industry heavily pollutes the domestic water and air (i.e., it exhibits a negative production externality). Now suppose the government decides to protect the infant industry with an import tariff. This action would, of course, stimulate domestic production of the good and also stimulate the positive learning effects for the economy. However, the domestic price of this good would rise, reducing domestic consumption. These higher prices would force consumers to substitute other products in consumption. Since the other industry's products are assumed to be substitutable, demand for that industry's goods will rise. The increase in demand would stimulate production of that good and, because of its negative externality, cause more pollution to the domestic environment. If the negative effects to the economy from additional pollution are greater than the positive learning effects, then the infant industry protection could reduce rather than improve national welfare.

The point of this example, however, is to demonstrate that in the presence of multiple distortions or imperfections in interconnected markets (i.e., in a general equilibrium model), the determination of optimal policies requires that one consider the intermarket effects. The optimal infant industry tariff must take into account the effects of the tariff on the polluting industry. Similarly, if the government wants to set an optimal environmental policy, it would need to account for the effects of the policy on the industry with the learning externality.

This simple example suggests a much more serious informational problem for the government. If the real economy has numerous market imperfections and distortions spread out among numerous industries that are interconnected through factor or goods market competition, then to determine the true optimal set of policies that would correct or reduce all the imperfections and distortions simultaneously would require the solution to a dynamic general equilibrium model that accurately describes the real economy not only today but also in all periods in the future. This type of model, or its solution, is simply not achievable today with any high degree of accuracy. Given the complexity, it seems unlikely that we would ever be capable of producing such a model.

The implication of this informational problem is that trade policy will always be like a shot in the dark. There is absolutely no way of knowing with a high degree of accuracy whether any policy will improve economic efficiency. This represents a serious blow to the case for government intervention in the form of trade policies. If the intention of government is to set trade policies that will improve economic

efficiency, then since it is impossible to know whether any policy would actually achieve that goal, it seems prudent to avoid the use of any such policy. Of course, the goal of government may not be to enhance economic efficiency, and that brings us to the last counterargument against selected protection.

Political Economy Issues: The Problem with Democratic Processes

In democratic societies, government representatives and officials are meant to carry out the wishes of the general public. As a result, decisions by the government are influenced by the people they represent. Indeed, one of the reasons “free speech” is so important in democratic societies is to assure that individuals can make their attitudes toward government policies known without fear of reproach. Individuals must be free to inform the government of which policies they approve and of which they disapprove if the government is truly to be a representative of the people. The process by which individuals inform the government of their preferred policies is generally known as lobbying.

In a sense, one could argue that lobbying can help eliminate some of the informational deficiencies faced by governments. After all, much of the information the government needs to make optimal policies is likely to be better known by its constituent firms and consumers. Lobbying offers a process through which information can be passed from those directly involved in production and consumption activities to the officials who determine policies. However, this process may turn out to be more of a problem than a solution.

One of the results of trade theory is that the implementation of trade policies will likely affect income distribution. In other words, all trade policies will generate income benefits to some groups of individuals and income losses to other groups. Another outcome, though, is that the benefits of protection would likely be concentrated—that is, the benefits would accrue to a relatively small group. The losses from protection, however, would likely be dispersed among a large group of individuals.

This outcome was seen clearly in the partial equilibrium analysis of a tariff. When a tariff is implemented, the beneficiaries would be the import-competing firms, which would face less competition for their product, and the government, which collects tariff revenue. The losses would accrue to the thousands or millions of consumers of the product in the domestic economy.

For example, consider a tariff on textile imports being considered by the government of a small, perfectly competitive economy. Theory shows that the sum of the benefits to the government and the firms will be exceeded by the losses to

consumers. In other words, national welfare would fall. Suppose the beneficiaries of protection are one hundred domestic textile firms that would each earn an additional \$1 million in profit as a result of the tariff. Suppose the government would earn \$50 million in additional tariff revenue. Thus the total benefits from the tariff would be \$150 million. Suppose consumers as a group would lose \$200 million, implying a net loss to the economy of \$50 million. However, suppose there are one hundred million consumers of the products. That implies that each individual consumer would lose only \$2.

Now, if the government bases its decision for protection on input from its constituents, then it is very likely that protection will be granted even though it is not in the nation's best interest. The reason is that textile firms would have an enormous incentive to lobby government officials in support of the policy. If each firm expects an extra \$1 million, it would make sense for the firms to hire a lobbying firm to help make their case before the government. The arguments to be used, of course, are (1) the industry will decline and be forced to lay off workers without protection, thus protection will create jobs; (2) the government will earn additional revenues that can be used for important social programs; and (3) the tax is on foreigners and is unlikely to affect domestic consumers (number 3 isn't correct, of course, but the argument is often used anyway). Consumers, on the other hand, have very little individual incentive to oppose the tariff. Even writing a letter to your representative is unlikely to be worth the \$2 potential gain. Plus, consumers would probably hear (if they hear anything at all) that the policy will create some jobs and may not affect the domestic price much anyway (after all, the tax is on foreigners).

The implication of this problem is that the lobbying process may not accurately relate to the government the relative costs and benefits that will arise due to the implementation of a trade policy. As a result, the government would likely implement policies that are in the special interests of those groups who stand to accrue the concentrated benefits from protection, even though the policy may generate net losses to the economy as a whole. Thus by maintaining a policy of free trade, an economy could avoid national efficiency losses that could arise with lobbying in a democratic system.

KEY TAKEAWAYS

- Selected protection may fail to raise national welfare when foreign country retaliations occur. This is a potential problem when many countries are large in international markets.
- Selected protection with a trade policy is typically second best. A purely domestic policy to correct the market imperfection is often the better, or first-best, policy.
- Selected protection requires detailed information in order to set the policy at a level that will assure an improvement in national welfare. Because the necessary information is often lacking, getting selected protection right may be impossible.
- Selected protection can be captured by special interests in the lobbying process in representative democracies, thereby making it less likely that maximum national welfare will be achieved.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a potentially welfare-reducing reaction to beggar-thy-neighbor trade policies.
 - b. The term used to describe the lowest-cost policy action that corrects for market distortions or imperfections.
 - c. The often overlooked deficiencies that affect the ability of government to set effective policies.
 - d. The term used to describe the process by which individuals inform the government of their preferred policies.
 - e. Economists applying the theory of the second best would argue that free trade is appropriate in spite of market imperfections because *these types of policies* are usually first best.

11.6 Free Trade as the “Pragmatically Optimal” Policy Choice

LEARNING OBJECTIVE

1. Understand the modern argument for free trade as a “pragmatically optimal” policy choice.

In summary, the economic argument in support of free trade is a sophisticated argument that is based on the interpretation of results from the full collection of trade theories developed over the past two or three centuries. These theories, taken as a group, do not show that free trade is the best policy for every individual in all situations. Instead, the theories show that there are valid arguments supporting both free trade and protectionism. To choose between the two requires a careful assessment of the pros and cons of each policy regime.

The argument for free trade presented here accepts the notion that free trade may not always be optimal in terms of maximizing economic efficiency. The argument also accepts that free trade may not generate the most preferred distribution of income. In theory, there are numerous cases in which selected protectionism can improve aggregate welfare or could establish a more equal distribution of income. Nevertheless, despite these theoretical possibilities, it remains unclear and perhaps unlikely that selected protectionism could achieve the intended results. First, in many instances, a trade policy is not the best way to achieve the intended improvement in economic efficiency, nor is it likely to be the most efficient way to achieve a more satisfactory distribution of income. Instead, purely domestic tax and subsidy policies dominate. Second, even when a trade policy is the best policy choice, the possibility of retaliations and the likelihood of informational deficiencies or distortions caused by the lobbying process are sufficiently large as to make the intended outcomes unknowable.

In addition, the process of information collection, lobbying, and policy implementation is a costly economic activity. Labor and capital resources are allocated by interest groups attempting to affect policies favorable to them. The government also must expend resources to gather information, to implement and administer policies, and to monitor the effectiveness of these policies. In the United States, the following agencies and groups devote at least some of their time to trade policy implementation: the Office of the United States Trade Representative, the International Trade Commission, the Department of Commerce, the Federal Trade Commission, the Department of Justice, the Congress, and the president, among

others. One must wonder whether the cost of this bureaucracy, together with the cost to the private sector to influence the decisions of the government, is worth it, especially when the outcomes are virtually unknowable.

Thus the conclusion reached by many economists is that while free trade may not be “technically optimal,” it remains “pragmatically optimal.” That is, given our informational deficiencies and the other problems inherent in any system of selected protectionism, free trade remains the policy most likely to produce the highest level of economic efficiency attainable.

KEY TAKEAWAY

- While free trade may not be “technically optimal,” it remains “pragmatically optimal”—that is, free trade remains the policy most likely to produce the highest level of economic efficiency that is practically attainable.

EXERCISE

1. **Jeopardy Question.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a policy that is relatively easy to implement and has strong positive characteristics but may not be best in all conceivable circumstances.

Chapter 12

Introductory Finance Issues: Current Patterns, Past History, and International Institutions

Economics is a social science whose purpose is to understand the workings of the real-world economy. An economy is something that no one person can observe in its entirety. We are all a part of the economy, we all buy and sell things daily, but we cannot observe all parts and aspects of an economy at any one time.

For this reason, economists build mathematical models, or theories, meant to describe different aspects of the real world. For some students, economics seems to be all about these models and theories, these abstract equations and diagrams. However, in actuality, economics is about the real world, the world we all live in.

For this reason, it is important in any economics course to describe the conditions in the real world before diving into the theory intended to explain them. In this case, in a textbook about international finance, it is very useful for a student to know some of the values of important macroeconomic variables, the trends in these variables over time, and the policy issues and controversies surrounding them.

This first chapter provides an overview of the real world with respect to international finance. It explains not only how things look now but also where we have been and why things changed along the way. It describes current economic conditions and past trends with respect to the most critical international macroeconomic indicators. In particular, it compares the most recent worldwide economic recession with past business cycle activity to put our current situation into perspective. The chapter also discusses important institutions and explains why they have been created.

With this overview about international finance in the real world in mind, a student can better understand why the theories and models in the later chapters are being developed. This chapter lays the groundwork for everything else that follows.

12.1 GDP, Unemployment, Inflation, and Government Budget Balances

LEARNING OBJECTIVE

1. Learn current values for several important macroeconomic indicators from a selected set of countries, including GDP, GDP per capita, unemployment rates, inflation rates, national budget balances, and national debts.

When someone reads the business and economics news it is common to see numerous values and figures used to describe the economic situation somewhere. For example, if you read a story about the Philippines you might read that the gross domestic product (GDP) is \$167 billion or that the GDP per person is \$3,500 per person, or that its unemployment rate is 7.1 percent and its inflation rate is now 2.8 percent. You might read that it has a government budget deficit of 3.7 percent of the GDP and a trade deficit of 5.2 percent of the GDP. But what does this all mean? How is someone supposed to interpret and understand whether the numbers indicate something good, bad, or neutral about the country?

One way to make judgments is to compare these numbers with other countries. To this end, the next few sections will present some recent data for a selected set of countries. Although memorizing these numbers is not so important, especially since they will all soon change, it is helpful to have an idea about what the values are for a few countries; or if not that, to know the approximate normal average for a particular variable. Thus it is useful to know that GDP per person ranges from about \$500 per year at the low end to about \$50,000 to \$75,000 per person at the high end. It is also useful to know that unemployment rates are normally less than 10 percent. So when you read that Zimbabwe recently had unemployment of 75 percent, a reader will know how unusually large that is. Once you also recognize that inflation rates are normally less than 10 percent, a rate of 10,000 percent will strike you as extraordinary.

Thus the values for some of these numbers will be helpful to make comparisons across countries today and to make comparisons over time for a particular country. Therefore, it can be very helpful to know the numbers for at least a few countries, or what may be deemed a set of reference countries. The countries in [Table 12.1 "GDP and GDP per Capita \(PPP in Billions of Dollars\), 2009"](#) were selected to provide a cross section of countries at different levels of economic development. Thus the United States, the European Union, and Japan represent the largest economies in

the world today. Meanwhile, countries like Brazil, Russia, India, and China are watched so closely today that they have acquired their own acronym: the BRIC countries. Finally, countries like Indonesia, Kenya, Ghana, and Burundi are among the poorest nations of the world. Note that in later tables other countries were substituted for the African countries because data are less difficult to obtain.

Gross Domestic Product around the World

Macroeconomics is the study of the interrelationships of aggregate economic variables. The most important of these, without question, is a country's **gross domestic product (GDP)**¹. GDP measures the total value of all goods and services produced by a country during a year. As such, it is a measure of the extent of economic activity in a country or the economic size of a country.

And because the consumption of goods and services is one way to measure an individual's economic well-being, it is easy to calculate the GDP per capita (i.e., per person) to indicate the average well-being of individuals in a country.

Details about how to measure and interpret GDP follow in subsequent chapters, but before doing so, it makes some sense to know a little about how economy size and GDP per person vary across countries around the world. Which are the biggest countries, and which are the smallest? Which countries provide more goods and services, on average, and which produce less? And how wide are the differences between countries? Table 12.1 "GDP and GDP per Capita (PPP in Billions of Dollars), 2009" provides recent information for a selected group of countries. Note that reported numbers are based on purchasing power parity (PPP), which is a better way to make cross-country comparisons and is explained later. A convenient source of the most recent comprehensive data from three sources (the International Monetary Fund [IMF], the World Bank, and the U.S. CIA) of GDP ([http://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(PPP\)](http://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP))) and GDP per person ([http://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(PPP\)_per_capita](http://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita)) is available at Wikipedia.

Table 12.1 GDP and GDP per Capita (PPP in Billions of Dollars), 2009

Country/Region (Rank)	GDP (Percentage in the World)	GDP per Capita (Rank)
World	68,997 (100)	10,433
European Union (1)	15,247 (22.1)	—
United States (2)	14,265 (20.7)	47,440 (6)
China (3)	7,916 (11.5)	5,970 (100)

1. Measures the total value of all goods and services produced by a country during a year.

Country/Region (Rank)	GDP (Percentage in the World)	GDP per Capita (Rank)
Japan (4)	4,354 (6.3)	34,116 (24)
India (5)	3,288 (4.8)	2,780 (130)
Russia (7)	2,260 (3.3)	15,948 (52)
Brazil (10)	1,981 (2.9)	10,466 (77)
South Korea (14)	1,342 (1.9)	27,692 (33)
Indonesia (17)	908 (1.3)	3,980 (121)
Kenya (82)	60 (nil)	1,712 (148)
Ghana (96)	34 (nil)	1,518 (152)
Burundi (158)	3 (nil)	390 (178)

Table 12.1 "GDP and GDP per Capita (PPP in Billions of Dollars), 2009" displays several things that are worth knowing. First, note that the United States and European Union each make up about one-fifth of the world economy; together the two are 42 percent. Throw Japan into the mix with the European Union and the United States and together they make up less than one-sixth of the world's population. However, these three developed nations produce almost one-half of the total world production. This is a testament to the high productivity in the developed regions of the world. It is also a testament to the low productivity in much of the rest of the world, where it takes another five billion people to produce the remaining half of the GDP.

The second thing worth recognizing is the wide dispersion of GDPs per capita across countries. The United States ranks sixth in the world at \$47,000 and is surpassed by several small countries like Singapore and Luxembourg and/or those with substantial oil and gas resources such as Brunei, Norway, and Qatar (not shown in Table 12.1 "GDP and GDP per Capita (PPP in Billions of Dollars), 2009"). Average GDP per capita in the world is just over \$10,000, and it is just as remarkable how far above the average some countries like the United States, Japan, and South Korea are as it is how far below the average other countries like China, India, Indonesia, and Kenya are. Perhaps most distressing is the situation of some countries like Burundi that has a GDP of only \$370 per person. (Other countries in a similar situation include Zimbabwe, Congo, Liberia, Sierra Leone, Niger, and Afghanistan.)

2. The percentage of the labor force that is currently not employed.
3. The rate of change in the general level of prices in an economy. Alternatively, the percentage change in the consumer price index over a period.

Unemployment and Inflation around the World

Two other key macroeconomic variables that are used as an indicator of the health of a national economy are the **unemployment rate**² and the **inflation rate**³. The

unemployment rate measures the percentage of the working population in a country who would like to be working but are currently unemployed. The lower the rate, the healthier the economy and vice versa. The inflation rate measures the annual rate of increase of the consumer price index (CPI). The CPI is a ratio that measures how much a set of goods costs this period relative to the cost of the same set of goods in some initial year. Thus if the CPI registers 107, it would cost \$107 (euros or whatever is the national currency) to buy the goods today, while it would have cost just \$100 to purchase the same goods in the initial period. This represents a 7 percent increase in average prices over the period, and if that period were a year, it would correspond to the annual inflation rate. In general, a relatively moderate inflation rate (about 0–4 percent) is deemed acceptable; however, if inflation is too high it usually contributes to a less effective functioning of an economy. Also, if inflation is negative, it is called deflation, and that can also contribute to an economic slowdown.

Table 12.2 Unemployment and Inflation Rates

Country/Region	Unemployment Rate (%)	Inflation Rate (%)
European Union	9.8 (Oct. 2009)	+0.5 (Nov. 2009)
United States	10.0 (Nov. 2009)	+1.8 (Nov. 2009)
China	9.2 (2008)	+0.6 (Nov. 2009)
Japan	5.1 (Oct. 2009)	-2.5 (Oct. 2009)
India	9.1 (2008)	+11.5 (Oct. 2009)
Russia	7.7 (Oct. 2009)	+9.1 (Nov. 2009)
Brazil	7.5 (Oct. 2009)	+4.2 (Nov. 2009)
South Korea	3.5 (Nov. 2009)	+2.4 (Nov. 2009)
Indonesia	8.1 (Feb. 2009)	+2.4 (Oct. 2009)
Spain	19.3 (Oct. 2009)	+0.3 (Nov. 2009)
South Africa	24.5 (Sep. 2009)	+5.8 (Nov. 2009)
Estonia	15.2 (Jul. 2009)	-2.1 (Nov. 2009)

Source: *Economist*, Weekly Indicators, December 17, 2009.

The unemployment rates and inflation rates in most countries are unusual in the reported period because of the economic crisis that hit the world in 2008. The immediate effect of the crisis was a drop in demand for many goods and services, a

contraction in GDP, and the loss of jobs for workers in many industries. In addition, prices were either stable or fell in many instances. When most economies of the world were booming several years earlier, a normal unemployment rate would have been 3 to 5 percent, while a normal inflation rate would stand at about 3 to 6 percent.

As [Table 12.2 "Unemployment and Inflation Rates"](#) shows, though, unemployment rates in most countries in 2009 are much higher than that, while inflation rates tend to be lower with several exceptions. In the United States, the unemployment rate has more than doubled, but in the European Union, unemployment was at a higher rate than the United States before the crisis hit, and so it has not risen quite as much. Several standouts in unemployment are Spain and South Africa. These are exceedingly high rates coming very close to the United States unemployment rate of 25 percent reached during the Great Depression in 1933.

India's inflation rate is the highest of the group listed but is not much different from inflation in India the year before of 10.4 percent. Russia's inflation this year has actually fallen from its rate last year of 13.2 percent. Japan and Estonia, two countries in the list, are reporting deflation this year. Japan had inflation of 1.7 percent in the previous year, whereas Estonia's rate had been 8 percent.

Government Budget Balances around the World

Another factor that is often considered in assessing the health of an economy is the state of the country's government budget. Governments collect tax revenue from individuals and businesses and use that money to finance the purchase of government provided goods and services. Some of the spending is on public goods such as national defense, health care, and police and fire protection. The government also transfers money from those better able to pay to others who are disadvantaged, such as welfare recipients or the elderly under social insurance programs.

Generally, if government were to collect more in tax revenue than it spent on programs and transfers, then it would be running a **government budget surplus**⁴ and there would be little cause for concern. However, many governments oftentimes tend to spend and transfer more than they collect in tax revenue. In this case, they run a **government budget deficit**⁵ that needs to be paid for or financed in some manner. There are two ways to cover a budget deficit. First, the government can issue Treasury bills and bonds and thus borrow money from the private market; second, the government can sometimes print additional money. If borrowing occurs, the funds become unavailable to finance private investment or consumption, and thus the situation represents a substitution of public spending

4. When total tax revenue by a government exceeds its total expenditures during a year.

5. When total expenditures by a government exceed its total tax revenue collections during a year.

for private spending. Borrowed funds must also be paid back with accrued interest, which implies that larger future taxes will have to be collected assuming that budget balance or a surplus is eventually achieved.

When governments borrow, they will issue Treasury bonds with varying maturities. Thus some will be paid back in one of two years, but others perhaps not for thirty years. In the meantime, the total outstanding balance of IOUs (i.e., I owe you) that the government must pay back in the future is called the **national debt**⁶. This debt is owed to whoever has purchased the Treasury bonds; for many countries, a substantial amount is purchased by domestic citizens, meaning that the country borrows from itself and thus must pay back its own citizens in the future. The national debt is often confused with a nation's international indebtedness to the rest of the world, which is known as its international investment position (defined in the next section).

Excessive borrowing by a government can cause economic difficulties. Sometimes private lenders worry that the government may become insolvent (i.e., unable to repay its debts) in the future. In this case, creditors may demand a higher interest rate to compensate for the higher perceived risk. To prevent that risk, governments sometimes revert to the printing of money to reduce borrowing needs. However, excessive money expansion is invariably inflationary and can cause long-term damage to the economy.

In Table 12.3 "Budget Balance and National Debt (Percentage of GDP), 2009", we present budget balances for a selected set of countries. Each is shown as a percentage of GDP, which gives a more accurate portrayal of the relative size. Although there is no absolute number above which a budget deficit or a national debt is unsustainable, budget deficits greater than 5 percent per year, those that are persistent over a long period, or a national debt greater than 50 percent of GDP tends to raise concerns among investors.

Table 12.3 Budget Balance and National Debt (Percentage of GDP), 2009

Country/Region	Budget Balance (%)	National Debt (%)
European Union	-6.5	—
United States	-11.9	37.5
China	-3.4	15.6
Japan	-7.7	172.1
India	-8.0	56.4

6. The total value of government debt outstanding at a point in time. Alternatively, the sum total face value of government bonds that have been sold but not yet redeemed.

Country/Region	Budget Balance (%)	National Debt (%)
Russia	-8.0	6.5
Brazil	-3.2	38.8
South Korea	-4.5	24.4
Indonesia	-2.6	29.3
Spain	-10.8	40.7
South Africa	-5.0	31.6
Estonia	-4.0	4.8

Source: *Economist*, Weekly Indicators, December 17, 2009, and the CIA World Factbook.

Note that all the budget balances for this selected set of countries are in deficit. For many countries, the deficits are very large, exceeding 10 percent in the U.S. and Spain. Although deficits for most countries are common, usually they are below 5 percent of the GDP. The reason for the higher deficits now is because most countries have increased their government spending to counteract the economic recession, while at the same time suffering a reduction in tax revenues also because of the recession. Thus budget deficits have ballooned around the world, though to differing degrees.

As budget deficits rise and as GDP falls due to the recession, national debts as a percent of GDP are also on the rise in most countries. In the United States, the national debt is still at a modest 37.5 percent, but recent projections suggest that in a few years it may quickly rise to 60 percent or 70 percent of the GDP. Note also that these figures subtract any debt issued by the government and purchased by another branch of the government. For example, in the United States for the past decade or more, the Social Security system has collected more in payroll taxes than it pays out in benefits. The surplus, known as the Social Security “trust fund,” is good because in the next few decades as the baby boom generation retires, the numbers of Social Security recipients is expected to balloon. But for now the surplus is used to purchase government Treasury bonds. In other words, the Social Security administration lends money to the rest of the government. Those loans currently sum to about 30 percent of GDP or somewhat over \$4 trillion. If we include these loans as a part of the national debt, the United States debt is now, according to the online national debt clock, more than \$12 trillion or about 85 percent of GDP. (This is larger than $37.5 + 30$ percent because the debt clock is an estimate of more recent figures and reflects the extremely large government budget deficit run in the previous year.)

Most other countries' debts are on a par with that of the U.S. with two notable exceptions. First, China and Russia's debts are fairly modest at only 15.6 percent and 6.5 percent of GDP, respectively. Second, Japan's national debt is an astounding 172 percent of GDP. It has arisen because the Japanese government has tried to extricate its economy from an economic funk by spending and borrowing over the past two decades.

KEY TAKEAWAYS

- GDP and GDP per capita are two of the most widely tracked indicators of both the size of national economies and an economy's capacity to provide for its citizens.
- In general, we consider an economy more successful if its GDP per capita is high, unemployment rate is low (3–5 percent), inflation rate is low and nonnegative (0–6 percent), government budget deficit is low (less than 5 percent of GDP) or in surplus, and its national debt is low (less than 25 percent).
 - The United States, as the largest national economy in the world, is a good reference point for comparing macroeconomic data.
 - The U.S. GDP in 2008 stood at just over \$14 trillion while per capita GDP stood at \$47,000. U.S. GDP made up just over 20 percent of world GDP in 2008.
 - The U.S. unemployment rate was unusually high at 10 percent in November 2009 while its inflation rate was very low at 1.8 percent.
 - The U.S. government budget deficit was at an unusually high level of 11.9 percent of GDP in 2009 while its international indebtedness made it a debtor nation in the amount of 37 percent of its GDP.
 - Several noteworthy statistics are presented in this section:
 - Average world GDP per person stands at around \$10,000 per person.
 - The GDP in the U.S. and most developed countries rises as high as \$50,000 per person.
 - The GDP in the poorest countries like Kenya, Ghana, and Burundi is less than \$2,000 per person per year.
 - U.S. unemployment has risen to a very high level of 10 percent; however, in Spain it sits over 19 percent, while in South Africa it is over 24 percent.
 - Inflation is relatively low in most countries but stands at over 9 percent in Russia and over 11 percent in India. In several countries like Japan and Estonia, deflation is occurring.

- Due to the world recession, budget deficits have grown larger in most countries, reaching almost 12 percent of GDP in the United States.
- The national debts of countries are also growing larger, and Japan's has grown to over 170 percent of GDP.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The approximate value of world GDP in 2008.
 - b. The approximate value of EU GDP in 2008.
 - c. The approximate value of U.S. GDP in 2008.
 - d. The approximate value of world GDP per capita in 2008.
 - e. The approximate value of EU GDP per capita in 2008.
 - f. The approximate value of U.S. GDP per capita in 2008.
 - g. The approximate value of South Africa’s unemployment rate in 2009.
 - h. The approximate value of India’s inflation rate in 2009.
 - i. The approximate value of the U.S. budget balance as a percentage of its GDP in 2009.
 - j. The approximate value of Japan’s national debt as a percentage of its GDP in 2009.
2. Use the information in [Table 12.1 "GDP and GDP per Capita \(PPP in Billions of Dollars\), 2009"](#) and [Table 12.3 "Budget Balance and National Debt \(Percentage of GDP\), 2009"](#) to calculate the dollar values of the government budget balance and the national debt for Japan, China, Russia, South Korea, and Indonesia.

12.2 Exchange Rate Regimes, Trade Balances, and Investment Positions

LEARNING OBJECTIVE

1. Learn current values for several important international macroeconomic indicators from a selected set of countries, including the trade balance, the international investment position, and exchange rate systems.

Countries interact with each other in two important ways: trade and investment. Trade encompasses the export and import of goods and services. Investment involves the borrowing and lending of money and the foreign ownership of property and stock within a country. The most important international macroeconomic variables, then, are the trade balance, which measures the difference between the total value of exports and the total value of imports, and the exchange rate, which measures the number of units of one currency that exchanges for one unit of another currency.

Exchange Rate Regimes

Because countries use different national currencies, international trade and investment requires an exchange of currency. To buy something in another country, one must first exchange one's national currency for another. Governments must decide not only how to issue its currency but how international transactions will be conducted. For example, under a traditional gold standard, a country sets a price for gold (say \$20 per ounce) and then issues currency such that the amount in circulation is equivalent to the value of gold held in reserve. In this way, money is "backed" by gold because individuals are allowed to convert currency to gold on demand.

Today's currencies are not backed by gold; instead most countries have a central bank that issues an amount of currency that will be adequate to maintain a vibrant growing economy with low inflation and low unemployment. A central bank's ability to achieve these goals is often limited, especially in turbulent economic times, and this makes monetary policy contentious in most countries.

One of the decisions a country must make with respect to its currency is whether to fix its exchange value and try to maintain it for an extended period, or whether to allow its value to float or fluctuate according to market conditions. Throughout

history, **fixed exchange rates**⁷ have been the norm, especially because of the long period that countries maintained a gold standard (with currency *fixed* to gold) and because of the fixed exchange rate system (called the Bretton Woods system) after World War II. However, since 1973, when the Bretton Woods system collapsed, countries have pursued a variety of different exchange rate mechanisms.

The International Monetary Fund (IMF), created to monitor and assist countries with international payments problems, maintains a list of country currency regimes. The list displays a wide variety of systems currently being used. The continuing existence of so much variety demonstrates that the key question, “Which is the most suitable currency system?” remains largely unanswered. Different countries have chosen differently. Later, this course will explain what is necessary to maintain a fixed exchange rate or **floating exchange rate**⁸ system and what are some of the pros and cons of each regime. For now, though, it is useful to recognize the varieties of regimes around the world.

Table 12.4 Exchange Rate Regimes

Country/Region	Regime
Euro Area	Single currency within; floating externally
United States	Float
China	Crawling peg
Japan	Float
India	Managed float
Russia	Fixed to composite
Brazil	Float
South Korea	Float
Indonesia	Managed float
Spain	Euro zone; fixed in the European Union; float externally
South Africa	Float
Estonia	Currency board

7. An exchange rate system in which the rate is fixed at a specified value by the government.

8. An exchange rate system in which the rate fluctuates, or floats, because it is determined by supply and demand in the private market.

Source: International Monetary Fund, De Facto Classification of Exchange Rate Regimes and Monetary Policy Framework, 2008.

Table 12.4 "Exchange Rate Regimes" shows the selected set of countries followed by a currency regime. Notice that many currencies—including the U.S. dollar, the Japanese yen, the Brazilian real, the South Korean won, and the South African rand—are independently floating, meaning that their exchange values are determined in the private market on the basis of supply and demand. Because supply and demand for currencies fluctuate over time, so do the exchange values, which is why the system is called *floating*.

Note that India and Indonesia are classified as “managed floating.” This means that the countries’ central banks will sometimes allow the currency to float freely, but at other times will nudge the exchange rate in one direction or another.

China is listed and maintaining a crawling peg, which means that the currency is essentially fixed except that the Chinese central bank is allowing its currency to appreciate slowly with respect to the U.S. dollar. In other words, the fixed rate itself is gradually but unpredictably adjusted.

Estonia is listed as having a currency board. This is a method of maintaining a fixed exchange rate by essentially eliminating the central bank in favor of a currency board that is mandated by law to follow procedures that will automatically keep its currency fixed in value.

Russia is listed as fixing to a composite currency. This means that instead of fixing to one other currency, such as the U.S. dollar or the euro, Russia fixes to a basket of currencies, also called a composite currency. The most common currency basket to fix to is the Special Drawing Rights (SDR), a composite currency issued by the IMF used for central bank transactions.

Finally, sixteen countries in the European Union are currently members of the euro area. Within this area, the countries have retired their own national currencies in favor of using a single currency, the euro. When all countries circulate the same currency, it is the ultimate in fixity, meaning they have fixed exchange rates among themselves because there is no need to exchange. However, with respect to other external currencies, like the U.S. dollar or the Japanese yen, the euro is allowed to float freely.

Trade Balances and International Investment Positions

9. Occurs when the value of exports exceeds the value of imports during a year.

One of the most widely monitored international statistics is a country’s trade balance. If the value of total exports from a country exceeds total imports, we say a country has a **trade surplus**⁹. However, if total imports exceed total exports, then

the country has a **trade deficit**¹⁰. Of course, if exports equal imports, then the country has balanced trade.

The terminology is unfortunate because it conveys a negative connotation to trade deficits, a positive connotation to trade surpluses, and perhaps an ideal connotation to trade balance. Later in the text, we will explain if or when these connotations are accurate and when they are inaccurate. Suffice it to say, for now, that sometimes trade deficits can be positive, trade surpluses can be negative, and trade balance could be immaterial.

Regardless, it is popular to decry large deficits as being a sign of danger for an economy, to hail large surpluses as a sign of strength and dominance, and to long for the fairness and justice that would arise if only the country could achieve balanced trade. What could be helpful at an early stage, before delving into the arguments and explanations, is to know how large the countries' trade deficits and surpluses are. A list of trade balances as a percentage of GDP for a selected set of countries is provided in [Table 12.5 "Trade Balances and International Investment Positions GDP, 2009"](#).

It is important to recognize that when a country runs a trade deficit, residents of the country purchase a larger amount of foreign products than foreign residents purchase from them. Those extra purchases are financed by the sale of domestic assets to foreigners. The asset sales may consist of property or businesses (a.k.a. investment), or it may involve the sale of IOUs (borrowing). In the former case, foreign investments entitle foreign owners to a stream of profits in the future. In the latter case, foreign loans entitle foreigners to a future repayment of principal and interest. In this way, trade and international investment are linked.

Because of these future profit takings and loan repayments, we say that a country with a deficit is becoming a debtor country. On the other hand, anytime a country runs a trade surplus, it is the domestic country that receives future profit and is owed repayments. In this case, we say a country running trade surpluses is becoming a creditor country. Nonetheless, trade deficits or surpluses only represent the debts or credits extended over a one-year period. If trade deficits continue year after year, then the total external debt to foreigners continues to grow larger. Likewise, if trade surpluses are run continually, then credits build up. However, if a deficit is run one year followed by an equivalent surplus the second year, rather than extending new credit to foreigners, the surplus instead will represent a repayment of the previous year's debt. Similarly, if a surplus is followed by an equivalent deficit, rather than incurring debt to foreigners, the deficit instead will represent foreign repayment of the previous year's credits.

10. Occurs when the value of imports exceeds the value of exports during a year.

All of this background is necessary to describe a country's **international investment position (IIP)**¹¹, which measures the total value of foreign assets held by domestic residents minus the total value of domestic assets held by foreigners. It corresponds roughly to the sum of a country's trade deficits and surpluses over its entire history. Thus if the value of a country's trade deficits over time exceeds the value of its trade surpluses, then its IIP will reflect a larger value of foreign ownership of domestic assets than domestic ownership of foreign assets and we would say the country is a net debtor. In contrast, if a country has greater trade surpluses than deficits over time, it will be a net creditor.

Note how this accounting is similar to that for the national debt. A country's national debt reflects the sum of the nation's government budget deficits and surpluses over time. If deficits exceed surpluses, as they often do, a country builds up a national debt. Once a debt is present, though, government surpluses act to retire some of that indebtedness.

The key differences between the two are that the national debt is public indebtedness to both domestic and foreign creditors whereas the international debt (i.e., the IIP) is both public and private indebtedness but only to foreign creditors. Thus repayment of the national debt sometimes represents a transfer between domestic citizens and so in the aggregate has no impact on the nation's wealth. However, repayment of international debt always represents a transfer of wealth from domestic to foreign citizens.

Table 12.5 Trade Balances and International Investment Positions GDP, 2009

Country/Region	Trade Balance (%)	Debtor (-)/Creditor (+) Position (%)
Euro Area	-0.9	-17.5
United States	-3.1	-24.4
China	+6.1	+35.1
Japan	+2.7	+50.4
India	-0.3	-6.8
Russia	+2.2	+15.1
Brazil	-0.8	-26.6
South Korea	+3.8	-57.9
Indonesia	+1.2	-31.4
Spain	-5.7	-83.6

11. A measure of the difference between the total value of domestic holdings of foreign assets and the value of foreign assets held in the domestic country.

Country/Region	Trade Balance (%)	Debtor (-)/Creditor (+) Position (%)
South Africa	-5.4	-4.1
Estonia	+5.8	-83.1

Sources: *Economist*, the IMF, and the China State Administration of Foreign Exchange. See *Economist*, Weekly Indicators, December 30, 2009; IMF Dissemination Standards Bulletin Board at <http://dsbb.imf.org/Applications/web/dsbbhome>; IMF GDP data from Wikipedia at http://en.wikipedia.org/wiki/List_of_countries_by_GDP_%28nominal%29; and China State Administration of Foreign Exchange at http://www.safe.gov.cn/model_safe_en/tjsj_en/tjsj_detail_en.jsp?ID=3030300000000000,18&id=4.

Table 12.5 "Trade Balances and International Investment Positions GDP, 2009" shows the most recent trade balances and international investment positions, both as a percentage of GDP, for a selected set of countries. One thing to note is that some of the selected countries are running trade deficits while others are running trade surpluses. Overall, the value of all exports in the world must equal the value of all imports, meaning that some countries' trade deficits must be matched with other countries' trade surpluses. Also, although there is no magic number dividing good from bad, most observers contend that a trade deficit over 5 percent of GDP is cause for concern and an international debt position over 50 percent is probably something to worry about. Any large international debt is likely to cause substantial declines in living standards for a country when it is paid back—or at least if it is paid back.

The fact that debts are sometimes defaulted on, meaning the borrower decides to walk away rather than repay, poses problems for large creditor nations. The more money one has lent to another, the more one relies on the good faith and effort of the borrower. There is an oft-quoted idiom used to describe this problem that goes, “If you owe me \$100, you have a problem, but if you owe me a million dollars, then I have a problem.” Consequently, international creditor countries may be in jeopardy if their credits exceed 30, 40, or 50 percent of GDP.

Note from the data that the United States is running a trade deficit of 3.1 percent of GDP, which is down markedly from about 6 percent a few years prior. The United States has also been running a trade deficit for more than the past thirty years and as a result has amassed a debt to the rest of the world larger than any other country, totaling about \$3.4 trillion or almost 25 percent of U.S. GDP. As such, the U.S. is referred to as the largest debtor nation in the world.

In stark contrast, during the past twenty-five or more years Japan has been running persistent trade surpluses. As a result, it has amassed over \$2.4 trillion of credits to the rest of the world or just over 50 percent of its GDP. It is by far the largest creditor country in the world. Close behind Japan is China, running trade surpluses for more than the past ten years and amassing over \$1.5 trillion of credits to other countries. That makes up 35 percent of its GDP and makes China a close second to Japan as a major creditor country. One other important creditor country is Russia, with over \$250 billion in credits outstanding or about 15 percent of its GDP.

Note that all three creditor nations are also running trade surpluses, meaning they are expending their creditor position by becoming even bigger lenders.

Like the United States, many other countries have been running persistent deficits over time and have amassed large international debts. The most sizeable are for Spain and Estonia, both over 80 percent of their GDPs. Note that Spain continues to run a trade deficit that will add to its international debt whereas Estonia is now running a trade surplus that means it is in the process of repaying its debt. South Korea and Indonesia are following a similar path as Estonia. In contrast, the Euro area, South Africa, and to a lesser degree Brazil and India are following the same path as the United States—running trade deficits that will add to their international debt.

KEY TAKEAWAYS

- Exchange rates and trade balances are two of the most widely tracked international macroeconomic indicators used to discern the health of an economy.
- Different countries pursue different exchange rate regimes, choosing variations of floating and fixed systems.
 - The United States, as the largest national economy in the world, is a good reference point for comparing international macroeconomic data.
 - The United States maintains an independently floating exchange rate, meaning that its value is determined on the private market.
 - The United States trade deficit is currently at 3.1 percent of GDP. This is down from 6 percent recently but is one of a string of deficits spanning over thirty years.
 - The U.S. international investment position stands at almost 25 percent of GDP, which by virtue of the U.S. economy size, makes the United States the largest debtor nation in the world.
 - Several other noteworthy statistics are presented in this section:
 - China maintains a crawling peg fixed exchange rate.
 - Russia fixes its currency to a composite currency while Estonia uses a currency board to maintain a fixed exchange rate.
 - Japan is the largest creditor country in the world, followed closely by China and more distantly by Russia.
 - Spain and Estonia are examples of countries that have serious international debt concerns, with external debts greater than 80 percent of their GDPs.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The de facto exchange rate regime implemented in China in 2008.
 - b. The de facto exchange rate regime implemented in the United States in 2008.
 - c. The de facto exchange rate regime implemented in Indonesia in 2008.
 - d. The de facto exchange rate regime implemented in Estonia in 2008.
 - e. The name for the exchange rate regime in which a fixed exchange rate is adjusted gradually and unpredictably.
 - f. The name for the exchange rate regime in which the exchange rate value is determined by supply and demand for currencies in the private marketplace.
 - g. The term for the measure of the total value of foreign assets held by domestic residents minus the total value of domestic assets held by foreigners.
 - h. This country was the largest creditor country in the world as of 2008.
2. Use the information in [Table 12.1 "GDP and GDP per Capita \(PPP in Billions of Dollars\), 2009"](#) and [Table 12.5 "Trade Balances and International Investment Positions GDP, 2009"](#) to calculate the dollar values of the trade balance and the international investment position for Japan, China, Russia, South Korea, and Indonesia.

12.3 Business Cycles: Economic Ups and Downs

LEARNING OBJECTIVES

1. Understand the distinctions between an economic recession and a depression.
2. Compare and contrast the current recession in the United States with previous economic downturns.
3. Recognize why the economic downturn in the 1930s is called the Great Depression.

In 2009 the world was in the midst of the largest economic downturn since the early 1980s. Economic production was falling and unemployment rising. International trade fell substantially everywhere in the world, while investment both domestically and internationally dried up.

The source of these problems was the bursting of a real estate bubble. Bubbles are fairly common in both real estate and stock markets. A bubble is described as a steady and persistent increase in prices in a market, in this case, in the real estate markets in the United States and abroad. When bubbles are developing, many market observers argue that the prices are reflective of true values despite a sharp and unexpected increase. These justifications fool many people into buying the products in the hope that the prices will continue to rise and generate a profit.

When the bubble bursts, the demand driving the price increases ceases and a large number of participants begin to sell off their product to realize their profit. When these occur, prices quickly plummet. The dramatic drop in real estate prices in the United States in 2007 and 2008 left many financial institutions near bankruptcy. These financial market instabilities finally spilled over into the **real sector**¹² (i.e., the sector where goods and services are produced), contributing to a world recession. As the current economic crisis unfolds, there have been many suggestions about similarities between this recession and the Great Depression in the 1930s. Indeed, it is common for people to say that this is the biggest economic downturn since the Great Depression. But is it?

To understand whether it is or not, it is useful to look at the kind of data used to measure recessions or depressions and to compare what has happened recently with what happened in the past. First, here are some definitions.

12. The sector of an economy where goods and services are produced.

An **economic recession**¹³ refers to a decline in a country's measured real gross domestic product (GDP) over a period usually coupled with an increasing aggregate unemployment rate. In other words, it refers to a decline in economic productive activity. How much of a decline is necessary before observers will begin to call it a recession is almost always arguable, although there are a few guidelines one can follow.

In the United States, it is typical to define a recession as two successive quarters of negative real GDP growth. This definition dates to the 1970s and is little more than a rule of thumb, but it is one that has become widely applied. A more official way to define a recession is to accept the pronouncements of the National Bureau of Economic Research (NBER). This group of professional economists looks at more factors than just GDP growth rates and will also make judgments about when a recession has begun and when one has ended. According to the NBER, the current recession began in December 2007 in the United States. However, it did not proclaim that until December 2008. Although the U.S. economy contracted in the fourth quarter of 2007, it grew in the first two quarters of 2008, meaning that it did not fulfill the two successive quarters rule. That wasn't satisfied until the last two quarters of 2008 both recorded a GDP contraction. As of January 2010, the U.S. economy continues in a recession according to the NBER. See the National Bureau of Economic Research, <http://www.nber.org/cycles.html>.

A very severe recession is referred to as a depression. How severe a recession has to be to be called a depression is also a matter of judgment. In fact in this regard there are no common rules of thumb or NBER pronouncements. Some recent suggestions in the press are that a depression is when output contracts by more than 10 percent or the recession lasts for more than two years. Based on the second definition and using NBER records dating the length of recessions, the United States experienced depressions from 1865 to 1867, 1873 to 1879, 1882 to 1885, 1910 to 1912, and 1929 to 1933. Using this definition, the current recession could be judged a depression if NBER dates the end of the contraction to a month after December 2009.

13. Refers to a decline in a country's measured real gross domestic product (GDP) over a period usually coupled with an increasing aggregate unemployment rate.
14. The term used to describe the cyclical pattern of economic expansions and contractions over time.

The opposite of a recession is an economic expansion or economic boom. Indeed, the NBER measures not only the contractions but the expansions as well because its primary purpose is to identify the U.S. economy's peaks and troughs (i.e., high points and low points). When moving from a peak to a trough the economy is in a recession, but when moving from a trough to a peak it is in an expansion or boom. The term used to describe all of these ups and downs over time is the **business cycle**¹⁴.

The business cycle has been a feature of economies since economic activity has been measured. The NBER identifies recessions going back to the 1800s with the earliest

listed in 1854. Overall, the NBER has classified thirty-four recessions since 1854 with an average duration of seventeen months. The longest recession was sixty-five months from 1873 to 1879, a contraction notable enough to be called the Great Depression until another one came along to usurp it in the 1930s. On the upside, the average economic expansion in the United States during this period lasted thirty-eight months, with the longest being 120 months from 1991 to 2001. Interestingly, since 1982 the United States has experienced three of its longest expansions segmented only by relatively mild recessions in 1991 and 2001. This had led some observers to proclaim, “The business cycle is dead.” Of course, that was until we headed into the current crisis. (See here for a complete listing of NBER recessions: <http://www.nber.org/cycles/cyclesmain.html>.)

The Recession of 2008–2009

Next, let’s take a look at how the GDP growth figures look recently and see how they compare with previous periods. First, growth rates refer to the percentage change in real GDP, which means that the effects of inflation have been eliminated. The rates are almost always reported in annual terms (meaning the growth rate over a year) even when the period is defined as one quarter. In the United States and most other countries, GDP growth rates are reported every quarter, and that rate represents how much GDP would grow during a year if the rate of increase proceeded at the same pace as the growth during that quarter. Alternatively, annual growth rates can be reported as the percentage change in real GDP from the beginning to the end of the calendar year (January 1 to December 31).

Table 12.6 "U.S. Real GDP Growth and Unemployment Rate, 2007–2009" presents the quarterly real GDP growth rates from the beginning of 2007 to the end of 2009 and the corresponding unemployment rate that existed during the middle month of each quarter. Note first that in 2007, GDP growth was a respectable 2 to 3 percent and unemployment was below 5 percent, signs of a healthy economy. However, by the first quarter in 2008, GDP became negative although unemployment remained low. Growth rebounded to positive territory in the second quarter of 2008 while at the same time unemployment began to rise rapidly. At this time, there was great confusion about whether the U.S. economy was stalling or whether it was experiencing a temporary slowdown. By late 2008, though, speculation about an impending recession came to an end. Three successive quarters of significant GDP decline occurred between the second quarter of 2008 and the end of the first quarter in 2009, while the unemployment rate began to skyrocket. By the middle of 2009, the decline of GDP subsided and reversed to positive territory by the third quarter. However, the unemployment rate continued to rise, though at a slower pace. What happens next is anyone’s guess, but to get a sense of the severity of this recession it is worth analyzing at least two past recessions: that of 1981 to 1982 and

the two that occurred in the 1930s, which together are known as the Great Depression.

Table 12.6 U.S. Real GDP Growth and Unemployment Rate, 2007–2009

Year.Quarter	Growth Rate (%)	Unemployment Rate (%)
2007.1	1.2	4.5
2007.2	3.2	4.5
2007.3	3.6	4.7
2007.4	2.1	4.7
2008.1	-0.7	4.8
2008.2	1.5	5.6
2008.3	-2.7	6.2
2008.4	-5.4	6.8
2009.1	-6.4	8.1
2009.2	-0.7	9.4
2009.3	2.2	9.7
2009.4	—	10.0

Sources: U.S. Bureau of Economic Analysis and U.S. Department of Labor.

The Recession of 1980–1982

At a glance the current recession most resembles the recessionary period from 1980 to 1982. The NBER declared two recessions during that period; the first lasting from January to July 1980 and the second lasting from July 1981 to November 1982. As can be seen in [Table 12.7 "U.S. Real GDP Growth and Unemployment Rate, 1980–1983"](#), GDP growth moved like a roller coaster ride. Coming off a sluggish period of **stagflation**¹⁵ in the mid-1970s, unemployment began somewhat higher at around 6 percent, while growth in 1979 (not shown) was less than 1 percent in several quarters. Then in the second quarter of 1980, GDP plummeted by almost 8 percent, which is much more severe than anything in the current recession. Note that the largest quarterly decrease in the U.S. GDP in the post-World War II era was -10.4 percent in the first quarter of 1958. In the same quarter, unemployment soared, rising over a percentage point in just three months. However, this contraction was short-lived since the GDP fell only another 0.7 percent in the third quarter and then

15. An economic situation characterized by low or negative GDP growth together with high inflation.

rebounded with substantial growth in the fourth quarter of 1980 and the first quarter of 1981. Notice that despite the very rapid increase in the GDP, unemployment hardly budged downward, remaining stubbornly fixed around 7.5 percent. The rapid expansion was short-lived, as the GDP tumbled again by over 3 percent in the second quarter of 1981 only to rise again by a healthy 5 percent in the third quarter. But once again, the economy plunged back into recession with substantial declines of 5 percent and over 6 percent for two successive quarters in the GDP in late 1981 and early 1982. Meanwhile, from mid-1981 until after the real rebound began in 1983, the unemployment rate continued to rise, reaching a peak of 10.8 percent in late 1982, the highest unemployment rate in the post–World War II period.

Table 12.7 U.S. Real GDP Growth and Unemployment Rate, 1980–1983

Year.Quarter	Growth Rate (%)	Unemployment Rate (%)
1980.1	+1.3	6.3
1980.2	-7.9	7.5
1980.3	-0.7	7.7
1980.4	+7.6	7.5
1981.1	+8.6	7.4
1981.2	-3.2	7.5
1981.3	+4.9	7.4
1981.4	-4.9	8.3
1982.1	-6.4	8.9
1982.2	+2.2	9.4
1982.3	-1.5	9.8
1982.4	+0.3	10.8
1983.1	+5.1	10.4
1983.2	+9.3	10.1
1983.3	+8.1	9.5
1983.4	+8.5	8.5

Sources: U.S. Bureau of Economics and Analysis (<http://www.bea.gov>) and U.S. Department of Labor (<http://www.dol.gov>).

If indeed the current recession turns out like the 1980 to 1983 episode, we might expect to see substantial swings in the GDP growth rates in future quarters in the United States. The ups and downs are analogous to a bicycle smoothly traversing along a smooth road when the rider suddenly hits a large obstruction. The obstruction jolts the bike to one side while the rider compensates to pull the bike upright. However, the compensation is often too much, and the bike swings rapidly to the opposite side. This too inspires an exaggerated response that pushes the bike again too quickly to the original side. In time, the rider regains his balance and directs the bike along a smooth trajectory. That is what we see in Table 12.7 "U.S. Real GDP Growth and Unemployment Rate, 1980–1983" of the last quarters in 1983, when rapid growth becomes persistent and unemployment finally begins to fall.

The other lesson from this comparison is to note how sluggishly unemployment seems to respond to a growing economy. In late 1980 and early 1981, unemployment didn't budge despite the rapid revival of economic growth. In 1983, it took almost a full year of very rapid GDP growth before the unemployment rate began to fall substantially. This slow response is why the unemployment rate is often called a lagging indicator of a recession; it responds only after the recession has already abated.

The Great Depression

During the current recession there have been many references to the Great Depression of the 1930s. One remark often heard is that this is the worst recession since the Great Depression. As we can see in Table 12.7 "U.S. Real GDP Growth and Unemployment Rate, 1980–1983", this is not quite accurate since the recession of the early 1980s can easily be said to have been worse than the current one...at least so far.

It is worth comparing numbers between the current period and the Depression years if only to learn how bad things really were during the 1930s. The Great Depression was a time that transformed attitudes and opinions around the world and can surely be credited with having established the necessary preconditions for the Second World War.

So let's take a look at how bad it really was. Once again, we'll consider the U.S. experience largely because the data are more readily available. However, it is worth remembering that all three of the economic downturns described here are notable in that they were worldwide in scope.

First of all, there is no quarterly data available for the 1930s as quarterly data in the United States first appeared in 1947. Indeed, there was no formal organized

collection of data in the 1930s for a variable such as GDP. Thus the numbers presented by the U.S. Bureau of Economic and Analysis (BEA) were constructed by piecing together available data.

A second thing to realize is that annual GDP growth rates tend to have much less variance than quarterly data. In other words, the highs are not as high and the lows not as low. This is because the annual data are averaging the growth rates over the four quarters. Also, sometimes economic downturns occur at the end of one year and the beginning of the next so that the calendar year growth may still be positive in both years. For example in 2008, even though GDP growth was negative in three of four quarters, the annual GDP growth that year somehow registered a +0.4 percent. Also in 1980, despite an almost 8 percent GDP drop in the second quarter, the annual GDP growth that year was -0.3 percent. The same is true for 1982, which registered two quarters of negative GDP growth at -6.4 percent and -1.5 percent but still the GDP fell annually at only -1.9 percent.

With this caveat in mind, the U.S. GDP growth rates for the 1930s are astounding. From 1930 to 1933, the United States registered annual growth rates of -8.6 percent, -6.5 percent, -13.1 percent, and -1.3 percent. The unemployment rate, which is estimated to have been around 3 percent in the 1920s, rose quickly in 1930 to 8.9 percent and continued to rise rapidly to a height of almost 25 percent in 1933. Although growth returned with vigor in 1934 and for another four years, the unemployment rate remained high and only slowly fell to 14.3 percent by 1937.

Table 12.8 U.S. Real GDP Growth and Unemployment Rate, 1930–1940

Year	Growth Rate (%)	Unemployment Rate (%)
1930	-8.6	8.9
1931	-6.5	15.9
1932	-13.1	23.6
1933	-1.3	24.9
1934	+10.9	21.7
1935	+8.9	20.1
1936	+13.0	17.0
1937	+5.1	14.3
1938	-3.4	19.0
1939	+8.1	17.2

Year	Growth Rate (%)	Unemployment Rate (%)
1940	+8.8	14.6

Sources: U.S. Bureau of Economics and Analysis and U.S. Department of Labor.

The NBER dated the first part of the Depression as having started in August 1929 and ending in March 1933. But a second wave came, another recession beginning in May 1937 and ending in June 1938. This caused GDP to fall by another 3.4 percent in 1938 while unemployment rose back above 15 percent for another two years.

The Great Depression is commonly used to refer to the economic crisis (or crises) that persisted for the entire decade of the 1930s, only truly coming to an end at the start of World War II. Even then it is worth mentioning that although GDP began to grow rapidly during World War II, with GDP growth from 1941 to 1943 at 17.1 percent, 18.5 percent, and 16.4 percent, respectively, and with U.S. unemployment falling to 1.2 percent in 1944, these data mask the fact that most of the extra production was for bullets and bombs and much of the most able part of the workforce was engaged in battle in the Atlantic and Pacific war theaters. In other words, the movement out of the Great Depression was associated with a national emergency rather than a more secure and rising standard of living.

Although the data presented only cover the United States, the Great Depression was a worldwide phenomenon. Without digging too deeply into the data or just by taking a quick look at Wikipedia's article on the Great Depression, it reveals the following: unemployment in 1932 peaked at 29 percent in Australia, 27 percent in Canada, and 30 percent in Germany. In some towns with specialized production in the United Kingdom, unemployment rose as high as 70 percent.

Needless to say, the Great Depression was indeed "great" in the sense that it was the worst economic downturn the world experienced in the twentieth century. In comparison, the current recession, which is coming to be known as the Great Recession, comes nowhere close to the severity of the Great Depression...at least for the moment (as of January 2010). A more accurate description of the current recession is that it is the worst since the 1980s in the United States. However, we should always be mindful of a second downturn as was seen in the late 1930s. Even after things begin to improve, economies can suffer secondary collapses. Hopefully, demands will soon rebound, production will sluggishly increase, and unemployment rates will begin to fall around the world. We will soon see.

KEY TAKEAWAYS

- The business cycle refers to the cyclical pattern of economic expansions and contractions. Business cycles have been a persistent occurrence in all modern economies.
- The current recession, sometimes called the Great Recession, is comparable in GDP decline and unemployment increases in the United States to the recessions in the early 1980s.
- The Great Depression of the 1930s displayed much greater decreases in GDP, showed much larger increases in unemployment, and lasted for a longer period than any economic downturn in the United States since then.
- The largest annual decrease in the U.S. GDP during the Great Depression was -13.1 percent while the highest unemployment rate was 24.9 percent.
- The largest quarterly decrease in the U.S. GDP during the current recession was -6.4 percent while the highest unemployment rate was 10.1 percent.
- The largest quarterly decrease in the U.S. GDP since World War II was -10.4 percent in the first quarter of 1958, while the highest unemployment rate was 10.8 percent in 1982.
- Of the thirty-four U.S. recessions since 1854 classified by the NBER, the longest was sixty-five months in the 1870s, whereas the average length was seventeen months.
- Of all the U.S. expansions since 1854 classified by the NBER, the longest was 120 months in the 1990s whereas the average length was thirty-eight months.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Approximately the worst U.S. quarterly economic growth performance between 2007 and 2009.
 - b. Approximately the worst U.S. quarterly economic growth performance between 1980 and 1983.
 - c. Approximately the worst U.S. annual economic growth performance between 1930 and 1940.
 - d. Approximately the best U.S. annual economic growth performance between 1930 and 1940.
 - e. Approximately the period of time generally known as the Great Depression.
 - f. Approximately the highest unemployment rate in the U.S. during the Great Depression.
 - g. Approximately the highest unemployment rate in Germany during the Great Depression.
 - h. Approximately the best U.S. annual economic growth performance in the midst of World War II.
 - i. The longest economic recession (in months) in the United States since 1854 as classified by the NBER.
 - j. The longest economic expansion (in months) in the United States since 1854 as classified by the NBER.
 - k. The term used to describe the cyclical pattern of economic expansions followed by economic contractions.

12.4 International Macroeconomic Institutions: The IMF and the World Bank

LEARNING OBJECTIVES

1. Learn about the origins of the World Bank and the International Monetary Fund.
2. Understand the purpose of the International Monetary Fund both during the fixed exchange rate regime from 1945 to 1973 and after 1973.

After the Great Depression, one of the things policymakers thought was important was to return the international economy to a system of fixed exchange rates. Before the Depression (i.e., in the 1920s and before), the world mostly maintained a gold standard. Under such a system, a country establishes two rules: first, it fixes its currency value to a weight of gold; second, it establishes convertibility between the currency and gold. This means that any individual holding the national currency is allowed to cash in the currency for its equivalent in gold upon demand.

In essence, the gold standard derives from a system in which gold itself was used as a currency in exchange. Since gold was sufficiently rare and because it was inherently valuable to people, it was an ideal substance to use as a store of value and a medium of exchange (as was silver). However, once trucking gold around became more difficult, it became easier for governments to issue paper currency but to back up that currency with gold on reserve. Thus currency in circulation was just a representation of actual gold in the government's vault, and if a person ever wished to see that actual gold, he or she could simply demand conversion.

There is much that can be said about how a gold standard operates, but that discussion is reserved for a later chapter. For our purposes here, it is sufficient to explain that the gold standard was a system of fixed exchange rates. For example, before the 1930s the United States fixed the dollar at \$20.67 per ounce of gold. During the same period, the United Kingdom fixed its currency at £4.24 per ounce. As a result of the gold-currency convertibility in both countries, this meant the dollar and pound were fixed to each other at a rate of \$4.875/£.

During the Depression years, most countries dropped off the gold standard because the loss of confidence threatened a complete conversion of currency to gold and the depletion of national gold reserves. But, as World War II drew to a close, experts

were assembled in Bretton Woods, New Hampshire, in the United States in 1944 to design a set of institutions that would help establish an effective international monetary system and to prevent some of the adjustment catastrophes that occurred after World War I. One such catastrophe occurred in Germany in 1922 to 1923 when a floating German currency resulted in one of the worst hyperinflations in modern history. Photos from that period show people with wheelbarrows full of money being used to make basic purchases. One way to prevent a reoccurrence was to establish a system of fixed exchange rates. As will be shown later, an important benefit of fixed exchange rates is the potential for such a system to prevent excessive inflation.

The **Bretton Woods Conference**¹⁶, more formally called the United Nations Monetary and Financial Conference, was held in July 1944. The purpose of the conference was to establish a set of institutions that would support international trade and investment and prevent some of the monetary instabilities that had plagued the world after World War I. The conference proposed three institutions, only two of which finally came into being.

The unsuccessful institution was the International Trade Organization (ITO), which was intended to promote the reduction of tariff barriers and to coordinate domestic policies so as to encourage a freer flow of goods between countries. Although a charter was drawn up for the ITO, the United States refused to sign onto it, fearing that it would subordinate too many of its domestic policies to international scrutiny. A subagreement of the ITO, the General Agreement on Tariffs and Trade (GATT), designed to promote multilateral tariff reductions, was established independently though.

The two successfully chartered institutions from the Bretton Woods Conference were the **International Bank for Reconstruction and Development (IBRD)**¹⁷ and the **International Monetary Fund (IMF)**¹⁸.

16. A monetary conference held in 1944 in Bretton Woods, New Hampshire, that laid the groundwork for the establishment of the World Bank and the IMF.

17. The core institution in the World Bank whose mission is to provide loans to assist postwar reconstruction and economic development.

18. The international organization created after World War II to oversee the Bretton Woods system of fixed exchange rates.

The IBRD is one component of a larger organization called the World Bank. Its purpose was to provide loans to countries to aid their reconstruction after World War II and to promote economic development. Much of its early efforts focused on reconstruction of the war-torn economies, but by the 1960s, its efforts were redirected to developing countries. The intent was to get countries back on their feet, economically speaking, as quickly as possible.

The second successfully chartered organization was the IMF. Its purpose was to monitor and maintain the stability of the fixed exchange rate system that was established. The system was not the revival of a gold standard but rather what is known as a gold-exchange standard. Under this system, the U.S. dollar was singled

out as the international reserve currency. Forty-four of the forty-five ratifying countries agreed to have their currency fixed to the dollar. The dollar in turn was fixed to gold at \$35 per ounce. The countries also agreed not to exchange officially held gold deposits for currency as had been the practice under the gold standard. However, countries agreed that officially held gold could be exchanged between central banks.

Another important requirement designed to facilitate the expansion of international trade was that countries agreed not to put any restrictions or controls on the exchange of currencies when that exchange was intended for transactions on the current account. In other words, individuals would be free to exchange one currency for another if they wanted to import goods from another country. However, currency controls or restrictions were allowed for transactions recorded on the financial accounts. This allowed countries to prevent foreign purchases of businesses and companies or to prevent foreign banks from lending or borrowing money. These types of restrictions are commonly known as **capital controls**¹⁹ (also, currency controls and/or exchange restrictions). These controls were allowed largely because it was believed they were needed to help maintain the stability of the fixed exchange rate system.

The way a fixed exchange system operates in general, and the way the Bretton Woods gold exchange standard operated in particular, is covered in detail in [Chapter 22 "Fixed Exchange Rates"](#). For now I will simply state without explanation that to maintain a credible fixed exchange rate system requires regular intervention in the foreign exchange markets by country central banks. Sometimes to maintain the fixed rate a country might need to sell a substantial amount of U.S. dollars that it is holding on reserve. These reserves are U.S. dollar holdings that had been purchased earlier, but sometimes a country can run what is called a balance of payments deficit—that is, run out of dollar reserves and threaten the stability of the fixed exchange rate system.

At the Bretton Woods Conference, participants anticipated that this scenario would be a common occurrence and decided that a “fund” be established to essentially “bail out” countries that suffered from balance of payments problems. That fund was the IMF.

19. Any type of government restriction on international financial, or capital, transactions. This may include taxes on financial transactions or regulations specifying whether foreign citizens can buy domestic assets or in what quantities.

The IMF was created to help stabilize exchange rates in the fixed exchange rate system. In particular, member countries contribute reserves to the IMF, which is then enabled to lend money to countries suffering balance of payments problems. With these temporary loans, countries can avoid devaluations of their currencies or other adjustments that can affect the confidence in the monetary system. Because the monies used by the IMF are contributions given by other countries in the group,

it is expected that once a balance of payments problem subsides that the money will be repaid. To assure repayment the IMF typically establishes conditions, known as conditionality, for the recipients of the loans. These conditions generally involve changes in monetary and fiscal policies intended to eliminate the original problems with the balance of payments in the first place.

The role of the IMF has changed more recently though. The fixed exchange rate system, under which the IMF is designed to operate, collapsed in 1973. Since that time, most of the major currencies in the world—including the U.S. dollar, the British pound, the Japanese yen, and many others—are floating. When a currency is allowed to float, its value is determined by supply and demand in the private market and there is no longer any need for a country's central bank to intervene. This in turn means that a country can no longer get into a balance of payments problem since that balance is automatically achieved with the adjustment in the exchange rate value. In essence the *raison d'être* of the IMF disappeared with the collapse of the Bretton Woods system.

Curiously, the IMF did not fall out of existence. Instead, it reinvented itself as a kind of lender of last resort to national governments. After 1973, the IMF used its “fund” to assist national governments that had international debt problems. For example, a major debt crisis developed in the early 1980s when national governments of Mexico, Brazil, Venezuela, Argentina, and eventually many other nations were unable to pay the interest on their external debt, or the money they borrowed from other countries. Many of these loans were either taken by the national governments or were guaranteed by the national governments. This crisis, known as the Third World debt crisis, threatened to bring down the international financial system as a number of major banks had significant exposure of foreign loans that were ultimately defaulted on. The IMF stepped in to provide “structural adjustment programs” in this instance. So the IMF not only loaned money for countries experiencing balance of payments crises but also now provided loans to countries that could not pay back their foreign creditors. And also, because the IMF wanted to get its money back (meaning the money contributed by the member nations), the structural adjustment loans came with strings attached: **IMF conditionality²⁰**.

20. Conditions that countries must promise to fulfill to receive a loan from the IMF. Conditions can include requirements to reduce government budget deficits, suggestions for privatization of state-owned enterprises, and strict controls on monetary expansion.

Since that time, the IMF has lent money to many countries suffering from external debt repayment problems. It stepped in to help Brazil and Argentina several times in the 1980s and later. It helped Mexico during the peso crisis in 1994. It assisted countries during the Asian currency crisis of 1997 and helped Russia one year later when the Asian contagion swept through.

Although the IMF has come under much criticism, especially because conditionality is viewed by some as excessively onerous, it is worth remembering that the IMF

makes loans, not grants. Thus it has the motivation to demand changes in policies that raise the chances of being repaid. These conditions have generally involved things like fiscal and monetary responsibility. That means reducing one's government budget deficit and curtailing the growth of the money supply. It also prescribed privatization that involves the sale or divestiture of state-owned enterprises. The free market orientation of these conditions came to be known as the **Washington Consensus**²¹.

Also mitigating the criticisms is the fact that the countries that participate in IMF programs are free to accept the loans, or not. To illustrate the alternative, Malaysia was one country that refused to participate in an IMF structural adjustment program during the Asian currency crisis and as a result did not have to succumb to any conditions. Thus it is harder to criticize the IMF's conditions when the countries themselves have volunteered to participate. In exchange for what were often tens of billions of dollars in loans, these countries were able to maintain their good standing in the international financial community.

Although controversial, the IMF has played a significant role in maintaining the international financial system even after the collapse of fixed exchange rates. One last issue worth discussing in this introduction is the issue of moral hazard. In the past thirty years or so, almost every time a country has run into difficulty repaying its external debt, the IMF has stepped in to assure continued repayment. That behavior sends a signal to international investors that the risk of lending abroad is reduced. After all, if the country gets into trouble the IMF will lend the country money and the foreign creditors will still get their money back. The moral hazard refers to the fact that lending institutions in the developed countries may view the IMF like an insurance policy and thus make much riskier loans than they would have otherwise. In this way, the IMF could be contributing to the problem of international financial crisis rather than merely being the institution that helps clean up the mess.

21. A set of free market-oriented policies often prescribed by the IMF as a part of its conditions for countries to receive a structural adjustment loan.

KEY TAKEAWAYS

- The World Bank and the IMF were proposed during the Bretton Woods Conference in 1944.
- The main purpose of the World Bank is to provide loans for postwar reconstruction and economic development for developing countries.
- The main purpose of the IMF was to monitor the international fixed exchange rate system and to provide temporary loans to countries suffering balance of payments problems.
- Since the breakup of the Bretton Woods fixed exchange rate system in 1973, the IMF has mostly assisted countries by making structural adjustment loans to those that have difficulty repaying international debts.
- The IMF conditionalities are the often-criticized conditions that the IMF places on foreign governments accepting their loans. The free-market orientation of these conditions is known as the Washington Consensus.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The name for the original division of the World Bank that describes its original purpose.
 - b. The name for the international institution that was designed to assist countries suffering from balance of payments problems.
 - c. The common name for the international institution whose primary function today is to make loans to countries to assist their economic development.
 - d. In the Bretton Woods system, these types of regulations were allowed for transactions recorded on the financial account.
 - e. This type of currency regime was implemented immediately after the collapse of the Bretton Woods system.
 - f. The term used for the conditions the IMF places on loans it makes to countries.
 - g. The term used for the type of loans made by the IMF to assist countries having difficulty making international debt repayments.
 - h. The term used to describe the standard free market package of conditions typically invoked by the IMF on loans it makes to countries.

Chapter 13

National Income and the Balance of Payments Accounts

The most important macroeconomic variable tracked by economists and the media is the gross domestic product (GDP). Whether it ought to be so important is another matter that is discussed in this chapter. But before that evaluation can occur, the GDP must be defined and interpreted. This chapter presents the national income identity, which defines the GDP. It also presents several other important national accounts, including the balance of payments, the twin-deficit identity, and the international investment position. These are the variables of prime concern in an international finance course.

13.1 National Income and Product Accounts

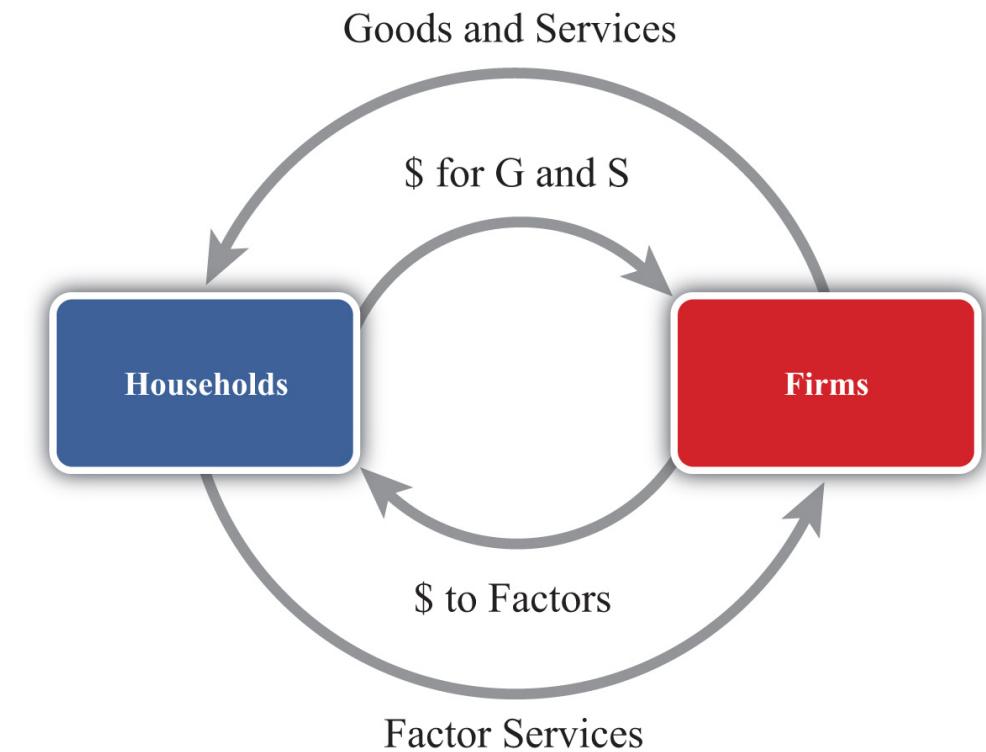
LEARNING OBJECTIVES

1. Define GDP and understand how it is used as a measure of economic well-being.
2. Recognize the limitations of GDP as a measure of well-being.

Many of the key aggregate variables used to describe an economy are presented in a country's National Income and Product Accounts (NIPA). National income represents the total amount of money that factors of production earn during the course of a year. This mainly includes payments of wages, rents, profits, and interest to workers and owners of capital and property. The national product refers to the value of output produced by an economy during the course of a year. National product, also called national output, represents the market value of all goods and services produced by firms in a country.

Because of the circular flow of money in exchange for goods and services in an economy, the value of aggregate output (the national product) should equal the value of aggregate income (national income). Consider the adjoining circular flow diagram, [Figure 13.1 "A Circular Flow Diagram"](#), describing a very simple economy. The economy is composed of two distinct groups: households and firms. Firms produce all the final goods and services in the economy using factor services (labor and capital) supplied by the households. The households, in turn, purchase the goods and services supplied by the firms. Thus goods and services move between the two groups in the counterclockwise direction. Exchanges are facilitated with the use of money for payments. Thus when firms sell goods and services, the households give the money to the firms in exchange. When the households supply labor and capital to firms, the firms give money to the households in exchange. Thus money flows between the two groups in a clockwise direction.

Figure 13.1 A Circular Flow Diagram



National product measures the monetary flow along the top part of the diagram—that is, the monetary value of goods and services produced by firms in the economy. National income measures the monetary flow along the bottom part of the diagram—that is, the monetary value of all factor services used in the production process. As long as there are no monetary leakages from the system, national income will equal national product.

The national product is commonly referred to as **gross domestic product (GDP)**¹. GDP is defined as the value of all final goods and services produced within the borders of a country during some period of time, usually a year. A few things are worth emphasizing about this definition.

First, GDP is measured in terms of the monetary (or dollar) value at which the items exchange in the market. Second, it measures only *final* goods and services as opposed to intermediate goods. Thus wheat sold by a farmer to a flour mill will not be directly included as part of GDP since the value of the wheat will be included in the value of the flour that the mill sells to the bakery. The value of the flour will in turn be included in the value of the bread sold to the grocery store. Finally, the value of the bread will be included in the price charged by the grocery when the product is finally purchased by the consumer. Only the final bread sale should be

1. Measures the total value of all goods and services produced by a country during a year.

included in GDP or else the intermediate values would overstate total production in the economy. Finally, GDP must be distinguished from another common measure of national output, **gross national product (GNP)**².

Briefly, GDP measures all production within the borders of the country regardless of who owns the factors used in the production process. GNP measures all production achieved by domestic factors of production regardless of where that production takes place. For example, if a U.S. resident owns a factory in Malaysia and earns profits on the operation of that factory, then those profits would be counted as production by a U.S. factory owner and thus would be included in the U.S. GNP. However, since that production took place beyond U.S. borders, it would not be counted as the U.S. GDP. Alternatively, if a Dutch resident owns a factory in the United States, then the fraction of that production that accrues to the Dutch owner would be counted as part of the U.S. GDP since the production took place in the United States. It would not be counted as part of the U.S. GNP, however, since the production was done by a foreign factor owner.

GDP is probably the most widely reported and closely monitored aggregate statistic. GDP is a measure of the size of an economy. It tells us the total amount of “stuff” the economy produces. Since most of us, as individuals, prefer to have more stuff rather than less, it is straightforward to extend this to the national economy to argue that the higher the GDP, the better off the nation. For this simple reason, statisticians track the growth rate of GDP. Rapid GDP growth is a sign of growing prosperity and economic strength. Falling GDP indicates a recession, and if GDP falls significantly, we call it an economic depression.

For a variety of reasons, GDP should be used only as a rough indicator of the prosperity or welfare of a nation. Indeed, many people contend that GDP is an inadequate measure of national prosperity. Below is a list of some of the reasons why GDP falls short as an indicator of national welfare.

2. A measure of national income that includes all production by citizens that occurs anywhere in the world. It is measured using the current account balance for exports and imports.

1. GDP only measures the amount of goods and services produced during the year. It does not measure the value of goods and services left over from previous years. For example, used cars, two-year-old computers, old furniture, old houses, and so on are all useful and provide welfare to individuals for years after they are produced. Yet the value of these items is only included in GDP in the year in which they are produced. National wealth, on the other hand, measures the value of all goods, services, and assets available in an economy at a point in time and is perhaps a better measure of national economic well-being than GDP.
2. GDP, by itself, fails to recognize the size of the population that it must support. If we want to use GDP to provide a rough estimate of the

average standard of living among individuals in the economy, then we ought to divide GDP by the population to get per capita GDP. This is often the way in which cross-country comparisons are made.

3. GDP gives no account of how the goods and services produced by the economy are distributed among members of the economy. One might prefer a lower GDP with a more equitable distribution to a higher GDP in which a small percentage of the population receives most of the product.
4. Measured GDP growth may overstate the growth of the standard of living since price level increases (inflation) would raise measured GDP. Thus even if the economy produces exactly the same amount of goods and services as the year before and prices of those goods rise, then GDP will rise as well. For this reason, real GDP is typically used to measure the growth rate of GDP. Real GDP divides nominal (or measured) GDP by the price level and is designed to eliminate some of the inflationary effects.
5. Sometimes, economies with high GDPs may also produce a large amount of negative production externalities. Pollution is one such negative externality. Thus one might prefer to have a lower GDP and less pollution than a higher GDP with more pollution. Some groups also argue that rapid GDP growth may involve severe depletion of natural resources, which may be unsustainable in the long run.
6. GDP often rises in the aftermath of natural disasters. Shortly after the Kobe earthquake in Japan in the 1990s, economists predicted that Japan's GDP would probably rise more rapidly. This is mostly because of the surge of construction activities required to rebuild the damaged buildings. This illustrates why GDP growth may not be indicative of a healthy economy in some circumstances.
7. GDP measures the value of production in the economy rather than consumption, which is more important for economic well-being. As will be shown later, national production and consumption are equal when a country's trade balance is zero; however, if a country has a trade deficit, then its national consumption will exceed its production. Ideally, because consumption is pleasurable while production often is not, we should use the measure of national consumption to measure economic well-being rather than GDP.

KEY TAKEAWAYS

- GDP is defined as the value of all final goods and services produced within the borders of a country during some period of time, usually a year.
- The following are several important weaknesses of GDP as a measure of economic well-being:
 - GDP measures income, not wealth, and wealth is a better measure of economic well-being.
 - GDP does not account for income distribution effects that may be important to economic well-being.
 - GDP measures “bads” like pollution as well as “goods.”
 - GDP measures production, not consumption, and consumption is more important to economic well-being.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term for the measure of national output occurring within the nation’s borders.
 - b. The term for the measure of national output that includes all production by domestic factors regardless of location.
 - c. Of *income* or *wealth*, this term better describes the gross domestic product (GDP).
 - d. Of *income* or *wealth*, this term better describes the gross national product (GNP).
 - e. The term used to describe the measure of GDP that takes account of price level changes or inflationary effects over time.
 - f. The term used to describe the measure of GDP that allows better income comparisons between countries that have different population sizes.
2. Many people argue that GDP is an inadequate measure of a nation’s economic well-being. List five reasons why this may be so.
3. GDP is used widely as an indicator of the success and economic well-being of the people of a nation. However, for many reasons it is not the perfect indicator. Briefly comment on the following statements related to this issue:
 - a. Domestic spending is a better indicator of standard of living than GDP.
 - b. National wealth is a better indicator of standard of living than GDP.

13.2 National Income or Product Identity

LEARNING OBJECTIVES

1. Identify the components of GDP defined in the national income identity.
2. Understand why imports are subtracted in the national income identity.

The national income or product identity describes the way in which the gross domestic product (GDP) is measured, as the sum of expenditures in various broad spending categories. The identity, shown below, says that GDP is the sum of personal consumption expenditures (C), private investment expenditures (I), government consumption expenditures (G), and expenditures on exports (EX) minus expenditures on imports (IM):

$$GDP = C + I + G + EX - IM.$$

Personal consumption expenditures (C), or “consumption” for short, include goods and services purchased by domestic residents. These are further subdivided into durable goods, commodities that can be stored and that have an average life of at least three years; nondurable goods, all other commodities that can be stored; and services, commodities that cannot be stored and are consumed at the place and time of purchase. Consumption also includes foreign goods and services purchased by domestic households.

Private domestic investment (I), or “investment” for short, includes expenditures by businesses on fixed investment and any changes in business inventories. Fixed investment, both residential and nonresidential, consists of expenditures on commodities that will be used in a production process for more than one year. It covers all investment by private businesses and by nonprofit institutions, regardless of whether the investment is owned by domestic residents or not. Nonresidential investment includes new construction, business purchases of new machinery, equipment, furniture, and vehicles from other domestic firms and from the rest of the world. Residential investment consists of private structures, improvements to existing units, and mobile homes. Note that this term does not include financial investments made by individuals or businesses. For example, one purchase of stock as an “investment” is not counted here.

Government expenditures include purchases of goods, services, and structures from domestic firms and from the rest of the world by federal, state, and local

government. This category includes compensation paid to government employees, tuition payments for higher education, and charges for medical care. Transfer payments, such as social insurance payments, government medical insurance payments, subsidies, and government aid are *not* included as a part of government expenditures.

Exports consist of goods and services that are sold to nonresidents.

Imports include goods and services purchased from the rest of the world.

The difference between exports and imports ($EX - IM$) is often referred to as net exports. Receipts and payments of factor income and transfer payments to the rest of the world (net) are excluded from net exports. Including these terms changes the trade balance definition and reclassifies national output as growth national product (GNP).

The Role of Imports in the National Income Identity

It is important to emphasize why imports are subtracted in the national income identity because it can lead to serious misinterpretations. First, one might infer (incorrectly) from the identity that imports are subtracted because they represent a cost to the economy. This argument often arises because of the typical political emphasis on jobs or employment. Thus higher imports imply that goods that might have been produced at home are now being produced abroad. This could represent an opportunity cost to the economy and justify subtracting imports in the identity. However, this argument is wrong.

The second misinterpretation that sometimes arises is to use the identity to suggest a relationship between imports and GDP growth. Thus it is common for economists to report that GDP grew at a slower than expected rate last quarter because imports rose faster than expected. The identity suggests this relationship because, obviously, if imports rise, GDP falls. However, this interpretation is also wrong.

The actual reason why imports are subtracted in the national income identity is because imports appear in the identity as hidden elements in consumption, investment, government, and exports. Thus imports must be subtracted to assure that only domestically produced goods are being counted. Consider the following details.

When consumption expenditures, investment expenditures, government expenditures, and exports are measured, they are measured without accounting for

where the purchased goods were actually made. Thus consumption expenditures (C) measures domestic expenditures on both domestically produced and foreign-produced goods. For example, if a U.S. resident buys a television imported from Korea, that purchase would be included in domestic consumption expenditures. Likewise, if a business purchases a microscope made in Germany, that purchase would be included in domestic investment. When the government buys foreign goods abroad to provide supplies for its foreign embassies, those purchases are included in government expenditures. Finally, if an intermediate product is imported, used to produce another good, and then exported, the value of the original imports will be included in the value of domestic exports.

This suggests that we could rewrite the national income identity in the following way:

$$GDP = (C_D + C_F) + (I_D + I_F) + (G_D + G_F) + (EX_D + EX_F) - IM,$$

where C_D represents consumption expenditures on domestically produced goods, C_F represents consumption expenditures on foreign-produced goods, I_D represents investment expenditures on domestically produced goods, I_F represents investment expenditures on foreign-produced goods, G_D represents government expenditures on domestically produced goods, G_F represents government expenditures on foreign-produced goods, EX_D represents export expenditures on domestically produced goods, and EX_F represents export expenditures on previously imported intermediate goods. Finally, we note that all imported goods are used in consumption, investment, or government or are ultimately exported, thus

$$IM = C_F + I_F + G_F + EX_F.$$

Plugging this expression into the identity above yields

$$GDP = C_D + I_D + G_D + EX_D$$

and indicates that GDP does not depend on imports at all.

The reason imports are subtracted in the standard national income identity is because they have already been included as part of consumption, investment, government spending, and exports. If imports were not subtracted, GDP would be overstated. Because of the way the variables are measured, the national income identity is written such that imports are added and then subtracted again.

This exercise should also clarify why the previously described misinterpretations were indeed wrong. Since imports do not affect the value of GDP in the first place, they cannot represent an opportunity cost, nor do they directly or necessarily influence the size of GDP growth.

KEY TAKEAWAYS

- GDP can be decomposed into consumption expenditures, investment expenditures, government expenditures, and exports of goods and services minus imports of goods and services.
- Investment in GDP identity measures physical investment, not financial investment.
- Government includes all levels of government and only expenditures on goods and services. Transfer payments are not included in the government term in the national income identity.
- Imports are subtracted in the national income identity because imported items are already measured as a part of consumption, investment and government expenditures, and as a component of exports. This means that imports have no direct impact on the level of GDP. The national income identity does not imply that rising imports cause falling GDP.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. A measure of the value of all capital equipment and services purchased during a year.
 - b. The term for the goods and services sold to residents of foreign countries.
 - c. The component of GDP that includes household purchases of durable goods, nondurable goods, and services.
 - d. The component of GDP that includes purchases by businesses for physical capital equipment used in the production process.
 - e. The government spending in the GDP identity does not count these types of government expenditures.
 - f. Of *true* or *false*, imported goods and services are counted once in the *C, I, G, or EX* terms of the GDP identity.

2. The national income identity says that gross domestic product is given by consumption expenditures, plus investment expenditures, plus government expenditures, plus exports, minus imports. In short, this is written as $GDP = C + I + G + EX - IM$.

Consider each of the following expenditures below. Indicate in which category or categories (*C, I, G, EX, or IM*) the item would be accounted for in the United States.

Product	Category
a. German resident purchase of a U.S.-made tennis racket	
b. U.S. firm purchase of a U.S.-made office copy machine	
c. Salaries to U.S. troops in Iraq	
d. School spending by county government	
e. U.S. household purchase of imported clothing	

3. What is the gross domestic product in a country whose goods and services balance is a \$300 billion deficit, consumption is \$900 billion, investment is \$300 billion, and government spending is \$500 billion?
4. Below are the economic data for the fictional country of Sandia. Write out the national income identity. Verify whether Sandia's data satisfy the identity.

TABLE 13.1 SANDIA'S ECONOMIC DATA
(BILLIONS OF DOLLARS)

Gross Domestic Product	400
Imports of Goods and Services	140
Investment Spending	20
Private Saving	30
Exports of Goods and Services	100
Government Transfers	40
Government Tax Revenues	140
Government Spending	140
Consumption Spending	280

13.3 U.S. National Income Statistics (2007–2008)

LEARNING OBJECTIVE

1. Learn the recent values for U.S. GDP and the relative shares of its major components.

To have a solid understanding of the international economy, it is useful to know the absolute and relative sizes of some key macroeconomic variables like the gross domestic product (GDP). For example, it is worthwhile to know that the U.S. economy is the largest in the world because its annual GDP is about \$14 trillion, not \$14 million or \$14 billion. It can also be useful to know about how much of an economy's output each year is consumed, invested, or purchased by the government. Although knowing that the U.S. government expenditures in 2008 were about \$2.9 trillion is not so important, knowing that government expenditures made up about 20 percent of GDP can be useful to know.

Table 13.2 "U.S. Gross Domestic Product (in Billions of Dollars)" contains U.S. statistics for the national income and product accounts for the years 2007 and 2008. The table provides the numerical breakdown of GDP not only into its broad components (C , I , G , etc.) but also into their major subcategories. For example, consumption expenditures are broken into three main subcategories: durable goods, nondurable goods, and services. The left-hand column indicates which value corresponds to the variables used in the identity.

Table 13.2 U.S. Gross Domestic Product (in Billions of Dollars)

		2007	2008	2008 (Percentage of GDP)
GDP	Gross domestic product	13,807.5	14,280.7	100.0
C	Personal consumption expenditures	9,710.2	10,058.5	70.4
	Durable goods	1,082.8	1,022.8	7.2
	Nondurable goods	2,833.0	2,966.9	20.8
	Services	5,794.4	6,068.9	42.5
I	Gross private domestic investment	2,134.0	2,004.1	14.0
	Nonresidential	1,503.8	1,556.2	10.9

		2007	2008	2008 (Percentage of GDP)
<i>G</i>	Structures	480.3	556.3	3.9
	Equipment and software	1,023.5	999.9	7.0
	Residential	630.2	487.8	3.4
	Change in business inventories	-3.6	-39.9	-0.0
<i>G</i>	Government consumption expenditures and gross investment	2,674.8	2,883.2	20.2
	Federal	979.3	1,071.2	7.5
	National defense	662.2	734.3	5.1
	Nondefense	317.1	336.9	2.4
	State and local	1,695.5	1,812.1	12.6
<i>EX</i>	Exports	1,662.4	1,867.8	13.1
	Goods	1,149.2	1,289.6	9.0
	Services	513.2	578.2	4.0
<i>IM</i>	Imports	2,370.2	2,533.0	17.7
	Goods	1,985.2	2,117.0	14.8
	Services	385.1	415.9	2.9

Source: Bureau of Economic Analysis, National Economic Accounts, Gross Domestic Product (GDP), at <http://www.bea.gov/national/nipaweb/Index.asp>.

There are a number of important things to recognize and remember about these numbers.

First, it is useful to know that U.S. GDP in 2008 was just over \$14 trillion (or \$14,000 billion). This is measured in 2008 prices and is referred to as nominal GDP. This number is useful to recall, first because it can be used in to judge relative country sizes if you happen to come across another country's GDP figure. The number will also be useful in comparison with U.S. GDP in the future. Thus if in 2020 you read that U.S. GDP is \$20 trillion, you'll be able to recall that back in 2008 it was just \$14 trillion. Also, note that between 2007 and 2008, the United States added over \$600 billion to GDP.

The next thing to note about the numbers is that consumption expenditures are the largest component of U.S. GDP, making up about 70 percent of output in 2008. That percentage is relatively constant over time, even as the economy moves between recessions and boom times (although it is up slightly from 68 percent in 1997). Notice also that services is the largest subcategory in consumption. This category includes health care, insurance, transportation, entertainment, and so on.

Gross private domestic investment, “investment” for short, accounted for just 14 percent of GDP in 2008. This figure is down from almost 17 percent just two years before and is reflective of the slide into the economic recession. As GDP began to fall at the end of 2008, prospects for future business opportunities also turned sour, and so investment spending also fell. As the recession continued into 2009, we can expect that number to fall even further the next year.

The investment component of GDP is often the target of considerable concern in the United States. Investment represents how much the country is adding to the capital stock. Since capital is an input into production, in general the more capital equipment available, the greater will be the national output. Thus investment spending is viewed as an indicator of future GDP growth. Perhaps the higher is investment, the faster the economy will grow in the future.

One concern about the U.S. investment level is that, as a percentage of GDP, it is lower than in many countries in Europe, especially in China and other Asian economies. In many European countries, it is above 20 percent of GDP. The investment figure is closer to 30 percent in Japan and over 35 percent in China. There was a fear among some observers, especially in the 1980s and early 1990s, that lower U.S. investment relative to the rest of the world would ultimately lead to slower growth. That this projection has not been borne out should indicate that higher investment is not sufficient by itself to assure higher growth.

Government expenditures on goods and services in the United States amounted to 20 percent of GDP in 2008. Due to the recession and the large government stimulus package in 2009, we can expect this number will rise considerably next year. Recall that this figure includes state, local, and federal spending but excludes transfer payments. When transfer payments are included, government spending plus transfers as a percentage of GDP exceeds 30 percent in the United States.

Two things are worth noting. First, the state and local spending is almost twice the level of federal spending. Second, most of the federal spending is on defense-related goods and services.

Exports in the United States accounted for 13 percent of GDP in 2008 (up from 10 percent in 2003) and are closing in on the \$2 trillion level. Imports into the United States are at \$2.5 trillion, amounting to almost 18 percent of GDP. In terms of the dollar value of trade, the United States is the largest importer and exporter of goods and services in the world. However, relative to many other countries, the United States trades less as a percentage of GDP.

KEY TAKEAWAYS

- U.S. GDP stands at just over \$14 trillion per year in 2008.
- U.S. consumption is about 70 percent of GDP; investment, 14 percent; government expenditures, 20 percent; exports, 13 percent; and imports, about 18 percent.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The approximate share of U.S. consumption as a share of U.S. GDP in 2008.
 - b. The approximate share of U.S. investment as a share of U.S. GDP in 2008.
 - c. The approximate share of U.S. government spending as a share of U.S. GDP in 2008.
 - d. The approximate share of U.S. exports of goods and services as a share of U.S. GDP in 2008.
 - e. The approximate share of U.S. imports of goods and services as a share of U.S. GDP in 2008.
 - f. This main category represents the largest share of GDP spending in the U.S. economy.

13.4 Balance of Payments Accounts: Definitions

LEARNING OBJECTIVES

1. Learn the variety of ways exports and imports are classified in the balance of payments accounts.
2. Understand the distinction between GDP and GNP.

The balance of payments accounts is a record of all international transactions that are undertaken between residents of one country and residents of other countries during the year. The accounts are divided into several subaccounts, the most important being the **current account**³ and the **financial account**⁴. The current account is often further subdivided into the merchandise trade account and the service account. These are each briefly defined in [Table 13.3 "Balance of Payments Accounts Summary"](#).

Table 13.3 Balance of Payments Accounts Summary

Current Account	Record of all international transactions for goods and services, income payments and receipts, and unilateral transfers . The current account is used in the national income identity for GNP.
Merchandise Trade Account ⁵	Record of all international transactions for goods only . Goods include physical items like autos, steel, food, clothes, appliances, furniture, etc.
Services Account ⁶	Record of all international transactions for services only . Services include transportation, insurance, hotel, restaurant, legal, consulting, etc.
Goods and Services Account ⁷	Record of all international transactions for goods and services only . The goods and services account is used in the national income identity for GDP.
Financial Account	Record of all international transactions for assets . Assets include bonds, Treasury bills, bank deposits, stocks, currency, real estate, etc.

- 3. A record of all international transactions for goods and services, income payments and receipts, and unilateral transfers.
- 4. A record of all international transactions for assets.
- 5. A record of all international transactions for goods only.
- 6. A record of all international transactions for services only.
- 7. A record of all international transactions for goods and services only.

The balance on each of these accounts is found by taking the difference between exports and imports.

Current Account

The current account (CA) balance is defined as $CA = EX^{G,S,IPR,UT} - IM^{G,S,IPR,UT}$ where the G,S,IPR,UT superscript is meant to include exports and imports of goods (G), services (S), income payments and receipts (IPR), and unilateral transfers (UT). If $CA > 0$, then exports of goods and services exceed imports and the country has a current account surplus. If $CA < 0$, then imports exceed exports and the country has a current account deficit.

Income payments represent the money earned (i.e., income) by foreign residents on their investments in the United States. For example, if a British company owns an office building in the United States and brings back to the United Kingdom a share of the profit earned there as a part of its income, then this is classified as an income payment on the current account of the balance of payments.

Income receipts represent the money earned by domestic residents on their investments abroad. For example, if a U.S. company owns an assembly plant in Costa Rica and brings back to the United States a share of the profit earned there as a part of its income, then this is classified as an income receipt on the current account of the balance of payments.

It may be helpful to think of income payments and receipts as payments for entrepreneurial services. For example, a British company running an office building is providing the management services and taking the risks associated with operating the property. In exchange for these services, the company is entitled to a stream of the profit that is earned. Thus income payments are classified as an import, the import of a service. Similarly, the U.S. company operating the assembly plant in Costa Rica is also providing entrepreneurial services for which it receives income. Since in this case the United States is exporting a service, income receipts are classified as a U.S. export.

Unilateral transfers represent payments that are made or received that do not have an offsetting product flow in the opposite direction. Normally, when a good is exported, for example, the good is exchanged for currency such that the value of the good and the value of the currency are equal. Thus there is an outflow and an inflow of equal value. An accountant would record both sides of this transaction, as will be seen in the next section. However, with a unilateral transfer, money flows out, but nothing comes back in exchange or vice versa. The primary examples of unilateral transfers are remittances and foreign aid. Remittances occur when a person in one country transfers money to a relative in another country and receives nothing in return. Foreign aid also involves a transfer, expecting nothing in return.

Merchandise Trade Balance

The merchandise trade balance (or goods balance) can be defined as $GB = EX^G - IM^G$, where we record only the export and import of merchandise goods. If $GB > 0$, the country would have a (merchandise) trade surplus. If $GB < 0$, the country has a trade deficit.

Services Balance

The service balance can be defined as $SB = EX^S - IM^S$, where we record only the export and import of services. If $SB > 0$, the country has a service surplus. If $SB < 0$, the country has a service deficit.

Goods and Services Balance

The goods and services balance (or goods balance) can be defined as $GSB = EX^{G\&S} - IM^{G\&S}$, where we record the export and import of both merchandise goods and services. If $GSB > 0$, the country would have a goods and services (G&S) surplus. If $GSB < 0$, the country has a G&S deficit. Note that sometimes people will refer to the difference $EX^{G\&S} - IM^{G\&S}$ as net exports. Often when this term is used the person is referencing the goods and services balance.

Here it is important to point out that when you hear a reference to a country's trade balance, it could mean the merchandise trade balance, or it could mean the goods and services balance, or it could even mean the current account balance.

Occasionally, one will hear trade deficit figures reported in the U.S. press followed by a comment that the deficit figures refer to the "broad" measure of trade between countries. In this case, the numbers reported refer to the current account deficit rather than the merchandise trade deficit. This usage is developing for a couple of reasons. First of all, at one time, around thirty years ago or more, there was very little international trade in services. At that time, it was common to report the merchandise trade balance since that accounted for most of the international trade. In the past decade or so, service trade has been growing much more rapidly than goods trade and it is now becoming a significant component of international trade. In the United States, service trade exceeds 30 percent of total trade. Thus a more complete record of a country's international trade is found in its current account balance rather than its merchandise trade account.

But there is a problem with reporting and calling it the current account deficit because most people don't know what the current account is. There is a greater chance that people will recognize the trade deficit (although most could probably

not define it either) than will recognize the current account deficit. Thus the alternative of choice among commentators is to call the current account deficit a trade deficit and then define it briefly as a “broad” measure of trade.

A simple solution would be to call the current account balance the “trade balance” since it is a record of all trade in goods and services and to call the merchandise trade balance the “merchandise goods balance,” or the “goods balance” for short. I will ascribe to this convention throughout this text in the hope that it might catch on.

GDP versus GNP

There are two well-known measures of the national income of a country: GDP and GNP. Both represent the total value of output in a country during a year, only measured in slightly different ways. It is worthwhile to understand the distinction between the two and what adjustments must be made to measure one or the other.

Conceptually, the gross domestic product (GDP) represents the value of all goods and services produced within the borders of the country. The gross national product (GNP) represents the value of all goods and services produced by domestic factors of production.

Thus production in the United States by a foreign-owned company is counted as a part of U.S. GDP since the productive activity took place within the U.S. borders, even though the income earned from that activity does not go to a U.S. citizen. Similarly, production by a U.S. company abroad will generate income for U.S. citizens, but that production does not count as a part of GDP since the productive activity generating that income occurred abroad. This production will count as a part of GNP though since the income goes to a U.S. citizen.

The way GDP versus GNP is measured is by including different items in the export and import terms. As noted above, GDP includes only exports and imports of goods and services, implying also that GDP excludes income payments and receipts and unilateral transfers. When these latter items are included in the national income identity and the current account balance is used for $EX - IM$, the national income variable becomes the GNP. Thus the GNP measure includes income payments and receipts and unilateral transfers. In so doing, GNP counts as additions to national income the profit made by U.S. citizens on its foreign operations (income receipts are added to GNP) and subtracts the profit made by foreign companies earning money on operations in the U.S. (income payments are subtracted).

To clarify, the national income identities for GDP and GNP are as follows:

$$GDP = C + I + G + EX^{G\&S} - IM^{G\&S}$$

and

$$GNP = C + I + G + EX^{G,S,IPR,UT} - IM^{G,S,IPR,UT}.$$

Financial Account Balance

Finally, the financial account balance can be defined as $KA = EX^A - IM^A$, where EX^A and IM^A refer to the export and import of assets, respectively. If $KA > 0$, then the country is exporting more assets than it is importing and it has a financial account surplus. If $KA < 0$, then the country has a financial account deficit.

The financial account records all international trade in assets. Assets represent all forms of ownership claims in things that have value. They include bonds, Treasury bills, stocks, mutual funds, bank deposits, real estate, currency, and other types of financial instruments. Perhaps a clearer way to describe exports of assets is to say that *domestic assets are sold to foreigners*, whereas *imports of assets mean foreign assets that are purchased by domestic residents*.

It is useful to differentiate between two different types of assets. First, some assets represent IOUs (i.e., I owe you). In the case of bonds, savings accounts, Treasury bills, and so on, the purchaser of the asset agrees to give money to the seller of the asset in return for an interest payment plus the return of the principal at some time in the future. These asset purchases represent borrowing and lending. When the U.S. government sells a Treasury bill (T-bill), for example, it is borrowing money from the purchaser of the T-bill and agrees to pay back the principal and interest in the future. The Treasury bill certificate, held by the purchaser of the asset, is an IOU, a promissory note to repay principal plus interest at a predetermined time in the future.

The second type of asset represents ownership shares in a business or property, which is held in the expectation that it will realize a positive rate of return in the future. Assets, such as common stock, give the purchaser an ownership share in a corporation and entitle the owner to a stream of dividend payments in the future if the company is profitable. The future sale of the stock may also generate a capital gain if the future sales price is higher than the purchase price. Similarly, real estate purchases—say, of an office building—entitle the owner to the future stream of rental payments by the tenants in the building. Owner-occupied real estate, although it does not generate a stream of rental payments, does generate a stream

of housing services for the occupant-owners. In either case, if real estate is sold later at a higher price, a capital gain on the investment will accrue.

An important distinction exists between assets classified as IOUs and assets consisting of ownership shares in a business or property. First of all, IOUs involve a contractual obligation to repay principal plus interest according to the terms of the contract or agreement. Failure to do so is referred to as a default on the part of the borrower and is likely to result in legal action to force repayment. Thus international asset purchases categorized as IOUs represent international borrowing and lending.

Ownership shares, on the other hand, carry no such obligation for repayment of the original investment and no guarantee that the asset will generate a positive rate of return. The risk is borne entirely by the purchaser of the asset. If the business is profitable, if numerous tenants can be found, or if real estate values rise over time, then the purchaser of the asset will make money. If the business is unprofitable, office space cannot be leased, or real estate values fall, then the purchaser will lose money. In the case of international transactions for ownership shares, there is no resulting international obligation for repayment.

KEY TAKEAWAYS

- The *trade balance* may describe a variety of different ways to account for the difference between exports and imports.
- The current account is the broadest measure of trade flows between countries encompassing goods, services, income payments and receipts, and unilateral transfers.
- The merchandise trade balance is a more narrow measure of trade between countries encompassing only traded goods.
- Net exports often refer to the balance on goods and services alone.
- GDP is a measure of national income that includes all production that occurs within the borders of a country. It is measured by using the goods and services balance for exports and imports.
- GNP is a measure of national income that includes all production by U.S. citizens that occurs anywhere in the world. It is measured by using the current account balance for exports and imports.
- The financial account balance measures all exports and imports of assets, which means foreign purchases of domestic assets and domestic purchases of foreign assets.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. A record of all international transactions for goods and services.
 - b. A record of all international transactions for assets.
 - c. The name of the balance of payments account that records transactions for goods.
 - d. The term used to describe the profit earned by domestic residents on their foreign business operations.
 - e. The term used to describe the profit earned by foreign residents on their domestic business operations.
 - f. The term used to describe remittances because they do not have a corresponding product flow to offset the money export or import.
 - g. Of *net importer* or *net exporter* of services, this describes a country that has more income payments than income receipts.
 - h. This measure of national output includes only the imports and exports of goods and services in its trade balance.
 - i. This measure of national output includes income payments and receipts in its trade balance.

13.5 Recording Transactions on the Balance of Payments

LEARNING OBJECTIVES

1. Learn how individual transactions between a foreign and domestic resident are recorded on the balance of payments accounts.
2. Learn the interrelationship between a country's current account balance and its financial account balance and how to interpret current account deficits and surpluses in terms of the associated financial flows.

In this section, we demonstrate how international transactions are recorded on the balance of payment accounts. The balance of payments accounts can be presented in ledger form with two columns. One column is used to record credit entries. The second column is used to record debit entries.

Almost every transaction involves an exchange between two individuals of two items believed to be of equal value. An exception is the case of unilateral transfers. These transfers include pension payments to domestic citizens living abroad, foreign aid, remittances, and other types of currency transfers that do not include an item on the reverse side being traded. Thus if one person exchanges \$20 for a baseball bat with another person, then the two items of equal value are the \$20 of currency and the baseball bat. The debit and credit columns in the ledger are used to record each side of every transaction. This means that every transaction must result in a credit and debit entry of equal value.

By convention, every credit entry has a “+” placed before it, while every debit entry has a “-” placed before it. The plus on the credit side generally means that money is being received in exchange for that item, while the minus on the debit side indicates a monetary payment for that item. This interpretation in the balance of payments accounts can be misleading, however, since in many international transactions, as when currencies are exchanged, money is involved on both sides of the transaction. There are two simple rules of thumb to help classify entries on the balance of payments:

1. Any time an item (good, service, or asset) is *exported* from a country, the value of that item is recorded as a credit entry on the balance of payments.
2. Any time an item is *imported* into a country, the value of that item is recorded as a debit entry on the balance of payments.

In the following examples, we will consider entries on the U.S. balance of payments accounts. Since it is a U.S. account, the values of all entries are denominated in U.S. dollars. Note that each transaction between a U.S. resident and a foreign resident would result in an entry on both the domestic and the foreign balance of payments accounts, but we will look at only one country's accounts.

Finally, we will classify entries in the balance of payments accounts into one of the two major subaccounts, the current account or the financial account. Any time an item in a transaction is a good or a service, the value of that item will be recorded in the current account. Any time an item in a transaction is an asset, the value of that item will be recorded in the financial account.

Note that in June 1999, what was previously called the “capital account” was renamed the “financial account” in the U.S. balance of payments. A capital account still exists but now includes only exchanges in nonproduced, nonfinancial assets. This category is very small, including such items as debt forgiveness and transfers by migrants. However, for some time, it will be common for individuals to use the term “capital account” to refer to the present “financial account.” So be warned.

A Simple Exchange Story

Consider two individuals, one a resident of the United States, the other a resident of Japan. We will follow them through a series of hypothetical transactions and look at how each of these transactions would be recorded on the balance of payments. The exercise will provide insight into the relationship between the current account and the financial account and give us a mechanism for interpreting trade deficits and surpluses.

Step 1: We begin by assuming that each individual wishes to purchase something in the other country. The U.S. resident wants to buy something in Japan and thus needs Japanese currency (yen) to make the purchase. The Japanese resident wants to buy something in the United States and thus needs U.S. currency (dollars) to make the purchase. Therefore, the first step in the story must involve an exchange of currencies.

So let's suppose the U.S. resident exchanges \$1,000 for ¥112,000 on the foreign exchange market at a spot exchange rate of 112 ¥/\$. The transaction can be recorded by noting the following:

1. The transaction involves an exchange of currency for currency. Since currency is an asset, both sides of the transaction are recorded on the financial account.
2. The currency exported is \$1,000 in U.S. currency. Hence, we have made a credit entry in the financial account in the table below. What matters is not whether the item leaves the country, but that the ownership changes from a U.S. resident to a foreign resident.
3. The currency imported into the country is the ¥112,000. We record this as a debit entry on the financial account and value it at the current exchange value, which is \$1,000 as noted in the table.

Step 1	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	0	
Financial Account	+1,000 (\$ currency)	-1,000 (¥ currency)

Step 2: Next, let's assume that the U.S. resident uses his ¥112,000 to purchase a camera from a store in Japan and then brings it back to the United States. Since the transaction is between the U.S. resident and the Japanese store owner, it is an international transaction and must be recorded on the balance of payments. The item exported in this case is the Japanese currency. We'll assume that there has been no change in the exchange rate and thus the currency is still valued at \$1,000. This is recorded as a credit entry on the financial account and labeled “¥ currency” in the table below. The item being imported into the United States is a camera. Since a camera is a merchandise good and is valued at ¥112,000 = \$1,000, the import is recorded as a debit entry on the current account in the table below.

Step 2	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	0	
Financial Account	+1,000 (¥ currency)	0

Step 3a: Next, let's assume that the Japanese resident uses his \$1,000 to purchase a computer from a store in the United States and then brings it back to Japan. The computer, valued at \$1,000, is being exported out of the United States and is considered a merchandise good. Therefore, a credit entry of \$1,000 is made in the following table on the current account and labeled as “computer.” The other side of the transaction is the \$1,000 of U.S. currency being given to the U.S. store owner by the Japanese resident. Since the currency, worth \$1,000, is being imported and is an

asset, a \$1,000 debit entry is made in the table on the financial account and labeled “\$ currency.”

Step 3a	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	+1,000 (computer)	
Financial Account	0	-1,000 (\$ currency)

Summary Statistics (after Steps 1, 2, and 3a)

We can construct summary statistics for the entries that have occurred so far by summing the debit and credit entries in each account and eliminating double entries. In the following table, we show all the transactions that have been recorded. The sum of credits in the current account is the \$1,000 computer. The sum of debits in the current account is the \$1,000 camera. On the financial account there are two credit entries of \$1,000, one representing U.S. currency and the other representing Japanese currency. There are two identical entries on the debit side. Since there is a U.S. currency debit and credit entry of equal value, this means that the net flow of currency is zero. The dollars that left the country came back in subsequent transactions. The same is true for Japanese currency. When reporting the summary statistics, the dollar and yen currency financial account entries would cancel, leaving a net export of assets equal to zero and the net inflow of assets equal to zero as well.

Summary 1, 2, 3a	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	+1,000 (computer)	
Financial Account	+1,000 (\$ currency), +1,000 (¥ currency)	-1,000 (\$ currency), -1,000 (¥ currency)

After cancellations, then, the summary balance of payments statistics would look as in the following table.

Summary 1, 2, 3a	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	+1,000 (computer)	
Financial Account	0	0

The current account balance is found by summing the credit and debit entries representing exports and imports, respectively. This corresponds to the difference between exports and imports of goods and services. In this example, the current account (or trade) balance is $CA = \$1,000 - \$1,000 = 0$. This means the trade account is balanced—exports equal imports.

The financial account balance is also found by summing the credit and debit entries. Since both entries are zero, the financial account balance is also zero.

Step 3b: Step 3b is meant to substitute for step 3a. In this case, we imagine that the Japanese resident decided to do something other than purchase a computer with the previously acquired \$1,000. Instead, let's suppose that the Japanese resident decides to save his money by investing in a U.S. savings bond. In this case, \$1,000 is paid to the U.S. government in return for a U.S. savings bond certificate (an IOU) that specifies the terms of the agreement (i.e., the period of the loan, interest rate, etc.). The transaction is recorded on the financial account as a credit entry of \$1,000 representing the savings bond that is exported from the country and a debit entry of \$1,000 of U.S. currency that is imported back into the country.

Step 3b	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	0	
Financial Account	+1,000 (U.S. savings bond)	-1,000 (\$ currency)

Summary Statistics (after Steps 1, 2, and 3b)

We can construct summary statistics assuming that steps 1, 2, and 3b have taken place. This is shown in the following table. The sum of credits in the current account in this case is zero since there are no exports of goods or services. The sum of debits in the current account is the \$1,000 camera.

On the financial account, there are three credit entries of \$1,000: one representing U.S. currency, the other representing Japanese currency, and the third representing the U.S. savings bond. There are two \$1,000 entries on the debit side: one representing U.S. currency and the other representing Japanese currency. Again, the dollar and yen currency financial account entries would cancel, leaving only a net export of assets equal to the \$1,000 savings bond. The net inflow of assets is equal to zero.

Summary 1, 2, 3b	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	0	
Financial Account	+1,000 (\$ currency), +1,000 (¥ currency), +1,000 (U.S. savings bond)	-1,000 (\$ currency), -1,000 (¥ currency)

After cancellations, the summary balance of payments statistics would look like the following table.

Summary 1, 2, 3b	U.S. Balance of Payments (\$)	
	Credits (+)	Debits (-)
Current Account	0	
Financial Account	+1,000 (U.S. savings bond)	0

The current account balance is found by summing the credit and debit entries representing exports and imports, respectively. This corresponds to the difference between exports and imports of goods and services. In this example, the current account (or trade) balance is $CA = \$0 - \$1,000 = -\$1,000$. This means there is a trade deficit of \$1,000. Imports of goods and services exceed exports of goods and services.

The financial account balance is also found by summing the credit and debit entries. In this example, the financial account balance is $KA = \$1,000 - \$0 = +\$1,000$. This means the financial account has a surplus of \$1,000. Exports of assets exceed imports of assets.

Important Lessons from the Exchange Story

The exercise above teaches a number of important lessons. The first lesson follows from the summary statistics, suggesting that the following relationship must hold true:

$$\text{current account balance} + \text{financial account balance} = 0.$$

In the first set of summary statistics (1, 2, 3a), both the current account and the financial account had a balance of zero. In the second example (1, 2, 3b), the current account had a deficit of \$1,000 while the financial account had a surplus of \$1,000.

This implies that anytime a country has a current account deficit, it *must* have a financial account surplus of equal value. When a country has a current account surplus, it *must* have a financial account deficit of equal value. And when a country has balanced trade (a balanced current account), then it *must* have balance on its financial account.

It is worth emphasizing that this relationship is *not* an economic theory. An economic theory could be right or it could be wrong. This relationship is an accounting identity. (That's why an identity symbol rather than an equal sign is typically used in the formula above.) An accounting identity is true by definition.

Of course, the identity is valid only if we use the *true* (or actual) current account and financial account balances. What countries report as their trade statistics are only the *measured* values for these trade balances, not necessarily the true values.

Statisticians and accountants attempt to measure international transactions as accurately as possible. Their objective is to record the true values or to measure trade and financial flows as accurately as possible. However, a quick look at any country's balance of payments statistics reveals that the balance on the current account plus the balance on the financial account rarely, if ever, sums to zero. The reason is not that the identity is wrong but rather that not all the international transactions on the balance of payments are accounted for properly. Measurement errors are common.

These errors are reported in a line in the balance of payments labeled "statistical discrepancy." The statistical discrepancy represents the amount that must be added or subtracted to force the measured current account balance and the measured financial account balance to zero. In other words, in terms of the measured balances on the balance of payments accounts, the following relationship will hold:

$$\text{current account balance} + \text{financial account balance} + \text{statistical discrepancy} = 0.$$

The second lesson from this example is that imbalances (deficits and surpluses) on the balance of payments accounts arise as a result of a series of mutually voluntary transactions in which equally valued items are traded for each other. This is an important point because it is often incorrectly interpreted that a trade deficit implies that unfair trade is taking place. After all, the logic goes, when imports

exceed exports, foreigners are *not* buying as many of our goods as we are buying of theirs. That's unequal exchange and that's unfair.

The story and logic are partially correct but incomplete. The logic of the argument focuses exclusively on trade in goods and services but ignores trade in assets. Thus it is true that when imports of goods exceed exports, we are buying more foreign goods and services than foreigners are buying of ours. However, at the same time, a current account deficit implies a financial account surplus. A financial account surplus, in turn, means that foreigners are buying more of our assets than we are buying of theirs. Thus when there is unequal exchange on the trade account, there must be equally opposite unequal exchange on the financial account. In the aggregate, imbalances on a current account, a trade account, or a financial account do not represent unequal exchanges between countries.

KEY TAKEAWAYS

- Every transaction between a domestic and foreign resident can be recorded as a debit and credit entry of equal value on the balance of payments accounts.
- All components of transactions that involve assets, including currency flows, are recorded on the financial account; all other items are recorded on the current account.
- All trade deficits on a country's current account implies an equally sized financial account surplus, while all trade surpluses implies an equally sized financial account deficit.
- In the aggregate, imbalances on a current account, a trade account, or a financial account do not represent unequal exchanges, or inequities, between countries.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The balance on a country’s financial account when its current account has a deficit of \$80 billion.
 - b. A country’s financial account balance when its trade balance is -\$60 billion, its service balance is +\$25 billion, and its unilateral transfer and income account has a surplus of +\$10 billion.
 - c. The international transactions for shares of stock in corporations (in excess of 10 percent of the company’s value) or for real estate.
 - d. Of *credit* or *debit*, this is how exports are recorded on the balance of payments.
 - e. Of *current account* or *financial account*, this is where an export of a clock will be recorded.
 - f. Of *current account* or *financial account*, this is where an import of currency from your aunt in Paraguay will be recorded.

2. Use the information below from the 1997 U.S. national income accounts to calculate the following. (Assume the balance on income and unilateral transfers was zero.)
 - a. Current account balance: _____
 - b. Merchandise trade balance: _____
 - c. Service balance: _____
 - d. Net income payments and receipts: _____
 - e. Goods and services balance: _____

TABLE 13.4 U.S. NATIONAL INCOME STATISTICS, 1997
(BILLIONS OF DOLLARS)

Gross Domestic Product	8,080
Exports of Goods and Services	934

Chapter 13 National Income and the Balance of Payments Accounts

Merchandise Exports	678
Income Receipts	257
Imports of Goods and Services	1,043
Merchandise Imports	877
Income Payments	244
Net Unilateral Transfers	-45

13.6 U.S. Balance of Payments Statistics (2008)

LEARNING OBJECTIVE

1. Learn the recent values for U.S. balance of payments statistics and the ways transactions are classified on both the current account and the financial account.

One of the most informative ways to learn about a country's balance of payments statistics is to take a careful look at them for a particular year. We will do that here for the U.S. balance of payments (U.S. BoP) statistics for 2008. Below we present an abbreviated version of the U.S. BoP statistics.

The line numbers refer to the line item on the complete Bureau of Economic Analysis (BEA) report. All debit entries have a minus sign, and all credit entries have a plus sign. A brief description of each line item is provided below where all values are rounded downward for easy reference with the table. To see the entries for every line or for more recent statistics, see the U.S. Department of Commerce, Bureau of Economic Analysis Web site, located at <http://www.bea.gov>.

Table 13.5 U.S. Balance of Payments, 2008 (Millions of Dollars Seasonally Adjusted)

Line Number	Category	Value (credits [+], debits [-])
Current Account		
1	Exports of goods, services, and income receipts	+2,591,233
3	Goods	+1,276,994
4	Services	+549,602
13	Income receipts on U.S. assets abroad	+761,593
14	Direct investment receipts	+370,747
15	Other private receipts	+385,940
16	U.S. government receipts	+4,906
18	Imports of goods, services, and income	-3,168,938
20	Goods	-2,117,245

Line Number	Category	Value (credits [+], debits [-])
21	Services	-405,287
30	Income payments on foreign assets in the United States	-636,043
31	Direct investment payments	-120,862
32	Other private payments	-349,871
33	U.S. government payments	-165,310
35	Unilateral transfers, net	-128,363
Capital Account		
39	Capital account transactions, net	+953
Financial Account		
40	U.S. assets abroad (increase/financial outflow [-])	-106
41	U.S. official reserve assets	-4,848
46	U.S. government assets	-529,615
50	U.S. private assets	+534,357
51	Direct investment	-332,012
52	Foreign securities	+60,761
53	U.S. claims reported by U.S. nonbanks	+372,229
54	U.S. claims reported by U.S. banks	+433,379
55	Foreign assets in the United States (increase/financial inflow [+])	+534,071
56	Foreign official assets in the United States	+487,021
63	Other foreign assets in the United States, net	+47,050
64	Direct investment	+319,737
65	U.S. Treasury securities	+196,619
66	U.S. securities other than T-bills	-126,737
67	U.S. currency	+29,187
68	U.S. liabilities reported by U.S. nonbanks	-45,167
69	U.S. liabilities reported by U.S. banks	-326,589
71	Statistical discrepancy (sum of above with sign reversed)	+200,055

Below we provide a brief description of each line item that appears on this abbreviated balance of payments record.

Current Account

Line 1, \$2.59 trillion, shows the value of all U.S. exports of goods, services, and income. This value is equal to the sum of lines 3, 4, and 13.

Line 3, \$1.27 trillion, shows exports of merchandise goods. This includes any physical items that leave the country.

Line 4, \$549 billion, shows exports of services to foreigners. This category includes travel services, passenger fares, royalties, license fees, insurance legal services, and other private services.

Line 13, \$761 billion, shows income receipts on U.S. assets abroad. This represents profits and interest earned by U.S. residents on investments in other countries. In a sense, these are payments for services rendered where the services include entrepreneurial services in the case of foreign-operated factories, or monetary services in the case of interest and dividend payments on foreign securities. This line is included in a measure of gross national product (GNP) since this income is accruing to U.S. factors of production. However, the line is excluded from a measure of gross domestic product (GDP) since production did not take place within the borders of the country. Income receipts are divided into four subcategories: direct investment receipts, other private receipts, U.S. government receipts, and compensation of employees.

Line 14, \$370 billion, shows direct investment receipts. This represents profit earned by U.S. companies on foreign direct investment (FDI), where FDI is defined as a greater than 10 percent ownership share in a foreign company. Note that this is not new investments but rather the profit and dividends earned on previous investments.

Line 15, \$385 billion, shows other private receipts. This category includes interest and profit earned by individuals, businesses, investment companies, mutual funds, pension plans, and so on. In effect, all private investment income that accrues on investments worth less than 10 percent of a company would be included here.

Line 16, \$4.9 billion, shows U.S. government income receipts. This refers to interest and other income earned by government investments abroad. Notice that this item is very small compared to the other two income categories.

Line 18, \$3.1 trillion, records imports of goods, services, and income. This value is equal to the sum of lines 20, 21, and 29.

Line 20, \$2.1 trillion, shows imports of merchandise goods. Notice that goods imports make up about two-thirds of total imports.

Line 21, \$405 billion, shows imports of services such as travel services, passenger fares, insurance, and so on.

Line 30, \$636 billion, shows income payments on foreign assets in the United States. This corresponds to income earned by foreigners who operate companies in the United States or income earned on other U.S.-based assets held by foreigners. This entry is further divided into four components: direct investment payments, other private payments, U.S. government payments, and compensation of employees.

Line 31, \$120 billion, records direct investment payments to foreigners in the United States. This represents profit earned on foreign direct investment by foreign residents' companies, where FDI is defined as a greater than 10 percent ownership share in a U.S. company. Note that this is not new investments but rather the profit and dividends earned on previous investments.

Line 32, \$349 billion, reports other private payments. This category includes interest and profit earned by individuals, businesses, investment companies, mutual funds, pension plans, and so on. In effect, all private investment income that accrues on investments worth less than 10 percent of a company would be included here.

Line 33, \$165 billion, records payments made by the U.S. government to foreigners. This item represents mostly interest payments on U.S. Treasury bills owned by foreigners.

Line 35, \$128 billion, records net unilateral transfers. These transfers refer to government grants to foreign nations, government pension payments, and private remittances to family and friends abroad. A debit entry here means that the net transfers are outbound, that is, more transfers are made from the U.S. to individuals abroad than are made in the reverse direction.

Capital Account

Line 39, \$953 million, represents net capital account transactions.

Financial Account

Line 40, \$106 million, shows the value of purchases of foreign assets by U.S. residents, hence it is referred to as a capital outflow. The line is the sum of U.S. official reserve assets (line 41), U.S. government assets (line 46), and U.S. private assets (line 50).

Line 41, \$4.8 billion, represents net U.S. official reserve transactions. Any purchases or sales of foreign currency in a foreign exchange intervention by the central bank would be recorded here. Since the item is a debit entry, it means that the U.S. central bank made net purchases of foreign assets (currencies) in 2008.

It is worth noting that this line is more important for a country maintaining a fixed exchange rate. To maintain a credible fixed exchange rate, central banks must periodically participate in the foreign exchange market. This line measures the extent of that participation and is sometimes referred to as the “balance of payments” in a fixed exchange rate system.

Line 46, \$529 billion, represents net purchases of assets by the U.S. government, though not by the Federal Reserve.

Line 50, \$534 billion, shows private purchases of foreign assets by U.S. residents. It is the primary component of total U.S. assets abroad. The item is composed of direct investment (line 51), foreign securities (line 52), U.S. claims reported by U.S. nonbanks (line 53), and U.S. claims reported by U.S. banks (line 54).

Line 51, \$332 billion, shows direct investment by U.S. residents abroad. It would include purchases of factories, stocks, and so on by U.S. businesses and affiliates in foreign countries as long as there is a controlling interest in excess of 10 percent voting share.

Line 52, \$60 billion, shows net purchases of foreign stocks and bonds by U.S. individuals and businesses when there is no controlling interest in the foreign company. Most purchases by U.S. mutual funds, pension funds, and insurance companies would be classified here.

Line 53, \$372 billion, shows U.S. resident purchases of foreign assets reported by nonbanks.

Line 54, \$433 billion, reports U.S. resident purchases of foreign assets reported by U.S. banks. This may include items like foreign currency denominated demand deposits held by U.S. businesses and individuals in U.S. banks.

Line 55, \$534 billion, shows the sum total of foreign assets in the United States. This item refers to all purchases of U.S. assets by foreign residents, thus, it is listed as a capital inflow. This line is composed of the sum of foreign official assets in the United States (line 56), and other foreign assets in the United States (line 63).

Line 56, \$487 billion, refers to purchases of U.S. assets by foreign governments or foreign central banks.

Line 63, \$47 billion, refers to all other foreign assets purchases of U.S. assets and is the main component of capital inflows. It is composed of direct investment (line 64), U.S. Treasury securities (line 65), U.S. securities other than T-bills (line 66), U.S. currency (line 67), U.S. liabilities reported by U.S. nonbanks (line 68), and U.S. liabilities reported by U.S. banks (line 69).

Line 64, \$319 billion, refers to purchases of U.S. factories and stocks when there is a greater than 10 percent ownership share.

Line 65, \$196 billion, shows total purchases of U.S. Treasury bills by foreigners. This corresponds to foreign loans to the U.S. government.

Line 66, \$126 billion, shows non-U.S. Treasury bill and nondirect investment purchases of stocks and bonds by foreigners.

Line 67, \$29 billion, a debit entry, represents U.S. currency that has been repatriated (net). Typically, this flow is a credit indicating an outflow of U.S. currency. Because of the expectation that the U.S. dollar will remain stable in value, it is often held by residents in inflationary countries to prevent the deterioration of purchasing power. It is estimated that over \$270 billion of U.S. currency circulates abroad and is used in exchange for foreign goods and services or simply held to store value. The value on line 67 represents only the amount that flowed back in 2007.

Line 68, \$45 billion, shows deposits and purchases of U.S. assets by foreigners reported by U.S. nonbanks.

Line 69, \$326 billion, reports deposits and purchases of U.S. assets by foreigners reported by U.S. banks. Thus if a foreign resident opens a checking account in a U.S. bank denominated in U.S. dollars, that value would be recorded here.

Line 71, \$200 billion, represents the statistical discrepancy. It is the sum of all the above items with the sign reversed. It is included to satisfy the accounting requirement that all debit entries be balanced by credit entries of equal value. Thus when the statistical discrepancy is included, the balance on the complete balance of payments is zero.

Summary Balances on the U.S. Balance of Payments (2008)

Table 13.6 "Balances on the U.S. Balance of Payments, 2008 (Millions of Dollars Seasonally Adjusted) (Credits [+], Debits [-])" reports a number of noteworthy balance of payments “balances” for 2008. In effect these subaccount balances allow us to identify net inflows or outflows of specific categories of goods, services, income, and assets.

Table 13.6 Balances on the U.S. Balance of Payments, 2008 (Millions of Dollars Seasonally Adjusted) (Credits [+], Debits [-])

Lines 1 + 18 + 35	Current account balance	-706, 068
Lines 3 + 20	Trade (goods) balance	-840, 251
Lines 4 + 21	Services balance	+144, 315
Lines 2 + 19	Goods and services balance	-695, 936
Lines 12 + 29	Investment income balance	+118, 231
Lines 40 + 55	Financial account balance	+533, 965
Line 71	Statistical discrepancy	+200,055

The sum of lines 1, 18, and 35 (i.e., exports of goods, services, and income; imports of goods, services, and income; and unilateral transfers [maintaining signs]) represents the current account (CA) balance. In 2008 in the United States, the CA balance was -706 billion dollars where the minus sign indicates a deficit. Thus the United States recorded a current account deficit of \$706 billion. Note that the current account balance is often reported as the “trade balance using a broad measure of international trade.”

Because unilateral transfers are relatively small and because investment income can be interpreted as payments for a service, it is common to say that a current account deficit means that imports of goods and services exceed exports of goods and services.

The sum of lines 3 and 20 (i.e., exports of goods and imports of goods) is known as the merchandise trade balance, or just trade balance for short. In 2008, the United States recorded a trade deficit of over \$840 billion. This means that the United States imported more physical goods than it exported.

The sum of lines 4 and 21, service exports and service imports, represents the service trade balance or just service balance. The table shows that the United States recorded a service surplus of over \$144 billion in 2008. In other words, the U.S. exports more services than it imports from the rest of the world.

The sum of lines 2 and 19, exports of goods and services and imports of goods and services, is a noteworthy trade balance because this difference is used in the national income identity for GDP. In contrast, the national income identity for GNP includes the current account balance instead. In 2008, the United States recorded a goods and services trade deficit of over \$695 billion.

The sum of lines 12 and 29, income receipts on U.S. assets abroad and income payments on foreign assets in the United States, represents the balance on investment income. In 2008, there was a recorded investment income surplus of over \$118 billion in the United States. This means that U.S. residents earned more on their investments abroad than foreigners earned on their investments in the United States.

The sum of lines 40 and 55, U.S. assets abroad and foreign assets in the United States, represents the financial account balance. In 2008, the United States recorded a financial account surplus of over \$533 billion. A surplus on capital account means that foreigners are buying more U.S. assets than U.S. residents are buying of foreign assets. These asset purchases, in part, represent international borrowing and lending. In this regard, a capital account surplus implies that the United States is borrowing money from the rest of the world.

Finally, line 70 records the 2008 U.S. statistical discrepancy as a \$200 billion credit entry. This implies that recorded debit entries on the balance of payments exceeded recorded credit entries. Thus an additional \$200 billion credit entry is needed to make the accounts balance. This is the largest statistical discrepancy recorded since the BEA records began in 1960.

The presence of a statistical discrepancy means that there are international transactions that have taken place but have not been recorded or accounted for properly. One might conclude that the size of the errors is \$200 billion, but this does not follow. The discrepancy only records the net effect. It is conceivable that \$400 billion of credit entries and \$200 billion of debit entries were missed. Or possibly, \$800 billion of debit entries and \$600 billion of credit entries were missed. In each case, the difference is \$200 billion dollars, but clearly the amount of error is substantially more in the latter case.

Based on the way the balance of payments data are collected, it seems likely that the primary source of the statistical discrepancy is on the capital account side rather than the current account side. This is because trade in goods, the primary component of the current account, is measured directly and completely by customs officials, while capital account data are acquired through surveys completed by major banks and financial institutions. This does not mean that errors cannot occur, however. Goods trade is tangible and thus is easier to monitor. Capital transactions, in contrast, can be accomplished electronically and invisibly and thus are more prone to measurement errors. Service and income transactions on the current account are also likely to exhibit the same difficulty in monitoring, implying that errors in the current account are more likely to arise in these subcategories.

KEY TAKEAWAYS

- The U.S. balance of payments records transactions on both the current and financial accounts concluding with several important balances.
- The United States had a current account deficit of \$706 billion in 2008.
- The U.S. had a merchandise trade deficit that was larger than its current account deficit at over \$840 billion in 2008.
- The U.S. had a financial account surplus of over \$533 billion.
- The statistical discrepancy at \$200 billion in 2008 demonstrates that all international transactions are not being recorded since the sum of the balance on the current account and the financial accounts does not equal zero.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The value of the statistical discrepancy if a country has a current account deficit of \$250 billion and a financial account surplus of \$230 billion.
 - b. The approximate value of the U.S. current account deficit in 2008.
 - c. The approximate value of the U.S. merchandise trade deficit in 2008.
 - d. Of *U.S. domestic residents* or *foreign residents*, this group profited more on its foreign investments because the United States ran a surplus on its investment income balance.
 - e. The approximate value of the U.S. financial account surplus in 2008.
 - f. The approximate value of the statistical discrepancy in the U.S. balance of payments in 2008.

13.7 The Twin-Deficit Identity

LEARNING OBJECTIVES

1. Learn the interrelationship between a country's government budget balance (deficit) and its current account balance (deficit).
2. Interpret the interrelationships of trade balances and budget balances in terms of the sources and uses of funds in the financial system.

One of the important relationships among aggregate economic variables is the so-called **twin-deficit identity**⁸, a term in reference to a country's government budget deficit and a simultaneous current account deficit. The name for this identity became commonplace during the 1980s and 1990s because at that time the United States experienced deficits in both of these accounts. Now, as we will see later, the identity will be a misnomer in many circumstances since there is no reason that "twin" deficits need to always appear together on these two national accounts. In fact, some countries will, at times, experience a deficit on one account and a surplus on the other. Also, at times, a country will experience a surplus on both accounts.

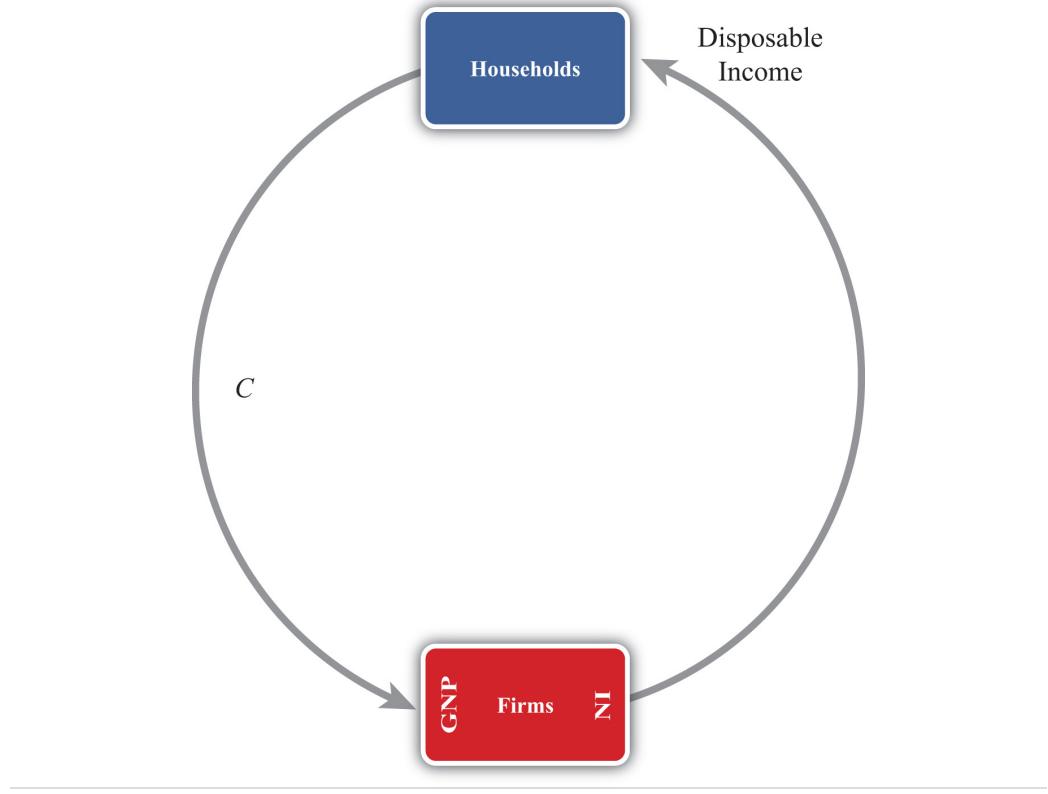
Thus a better title to this section would be "The Relationship between a Country's Government Budget Deficit and Its Current Account Deficit." However, since 2004, the United States finds itself back in the twin-deficit scenario, and since "twin-deficit identity" rolls off the tongue much more easily, we will stick to this title.

To understand this identity it will be helpful to take a much more careful look at the national income identity. This time I will build up the identity in a stepwise fashion using a circular flow diagram to better visualize the flows of money within an economy. A circular flow diagram is typically one of the first principles shown to students in an introductory macroeconomics class. Its purpose is to show the flow of money between the major players (or agents) within an economy. Circular flow diagrams can be either simple or complex depending on how many agents one introduces into the system and how finely one wishes to break down the monetary flows.

8. Shows the relationship between a country's government budget deficit and a country's current account deficit. It says that the sum of net private saving ($S_p - I$) and the current account deficit must equal the government budget deficit.

Circular Flow: Version 1

Figure 13.2 The Simplest Circular Flow

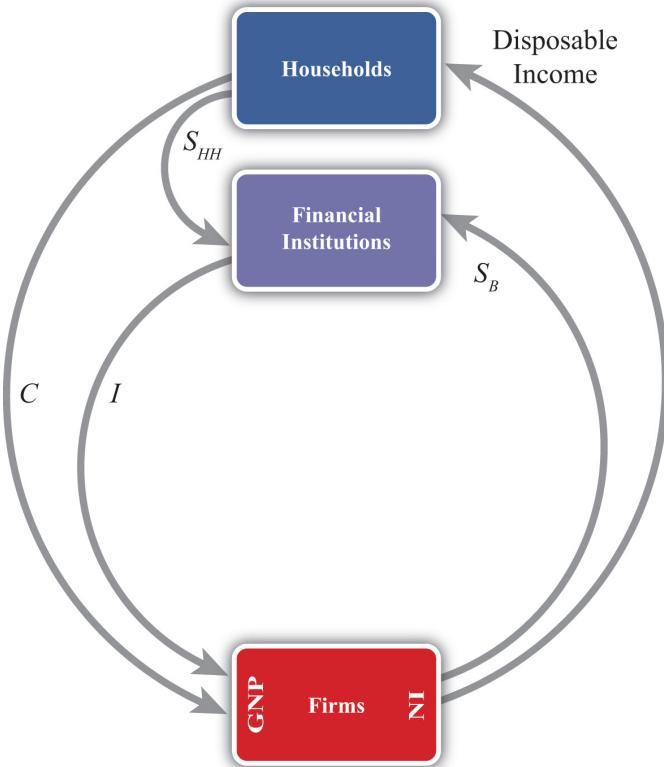


The simplest version of a circular flow diagram considers an economy consisting of two agents: households and firms. We imagine that firms produce goods and services using labor as an input.

The flow of money is shown in [Figure 13.2 "The Simplest Circular Flow"](#). The C arrow represents the dollar value of consumption expenditures made by households to purchase the goods and services produced and sold by firms. (The goods and services flow could be represented by an arrow in the opposite direction to C , but we leave that out for simplicity.) Since we assume in this case that there are only households buying goods, all GNP consists of C . The money that flows to firms from sales of consumption goods is given to the workers in exchange for their labor services. This monetary flow is represented by the arrow labeled “disposable income.” Disposable income is all the money households have to spend, which in this case is equal to the national income (NI).

Note especially that we use GNP rather than GDP as our measure of national income so that flows with the rest of the world later are properly defined.

Circular Flow: Version 2

Figure 13.3 *The Circular Flow Adding Financial Institutions*

The circular flow can be extended one step by adding financial institutions in [Figure 13.3 "The Circular Flow Adding Financial Institutions"](#). Financial institutions represent any company that facilitates borrowing and lending; the prime example is a bank. However, they may also include investment companies, pension funds, and mutual funds. The presence of financial institutions allows some money to be diverted from the consumption flow. In [Figure 13.3 "The Circular Flow Adding Financial Institutions"](#), these diversions are represented by S_{HH} , representing household savings and S_B , representing business saving. Some of the revenue earned by firms is not actually given out to workers in the form of wages. Instead some money is “retained” in the form of profit and excess earnings. These retained earnings are generally used to purchase investment goods to help an industry replace worn-out capital equipment and to add new capital. Much of these retained earnings may be used directly to purchase new capital equipment, although some of it will be saved by depositing it in a financial institution. For simplicity we will imagine that all such business saving flows through the financial system, hence the S_B arrow. In addition, households generally hold back some of their income from spending and deposit it into pension plans, savings accounts, and so on. Thus we

include the arrow from households. The easiest way to think of the diagram is to imagine that financial institutions take deposits from firms and households and then lend out the money to finance investment spending, I . With some exceptions, this is the way it will often work. One notable exception is that some of the money lent by banks is often used to finance consumption rather than investment. This occurs whenever households finance consumption spending using a credit card. However, we can avoid this complication by defining S_{HH} as being “net” savings, where the net means “after subtracting household borrowing.” With this definition in mind, it should be clear that S_{HH} can be negative—that is, its flow reversed—if household borrowing exceeds household saving.

We can now identify several important relationships. The first one relates to an important decision made by households. They choose how much of their disposable income should be spent on consumption and how much should be saved. You may recall from previous courses that the fraction of income spent on consumption goods (from an extra dollar of income) is called the **marginal propensity to consume**⁹, while the fraction of income saved is called the marginal propensity to save.

A second relationship is shown on the left side of the Firms box. This indicates that GNP is equal to the sum of C and I . This version of the national income identity would only be valid if there were no government sector and no trade with the rest of the world.

A third important relationship is shown by noting the flow of money in and out of the financial sector. There we see two arrows flowing in (i.e., S_{HH} and S_B) and one flow outward (i.e., I). This leads to the identity

$$S_{HH} + S_B = I,$$

indicating that the sum of household and business saving equals investment. A more common simplification of this relationship is shown by noting the following:

$$S_P = S_{HH} + S_B,$$

where S_P is called private saving. Thus private saving equals the sum of household saving and business saving. This will simplify the above identity to

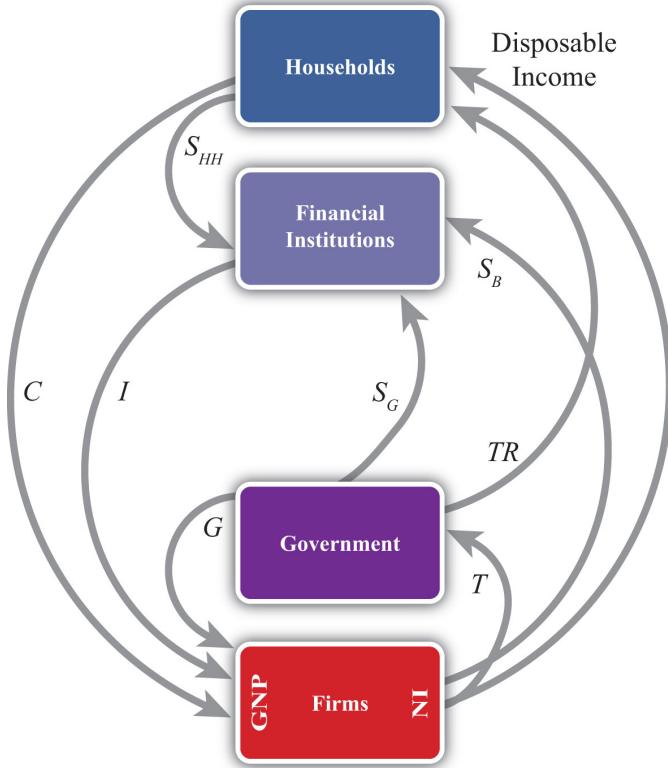
9. The additional amount of aggregate consumption that arises from an additional dollar of national income.

$$S_P = I,$$

or simply, private saving equals investment. Note that the term “private” is used here to distinguish it from government (or public sector) saving, which we’ll include next.

Circular Flow: Version 3

Figure 13.4 *The Circular Flow Adding Government*



Next, let's add in the government sector in [Figure 13.4 "The Circular Flow Adding Government"](#). The government is shown both to take money out of the circular flow and to inject money back in. Money is withdrawn first in the form of taxes (T). In the adjoining diagram, taxes are represented as a flow of money directly from firms, as if it is entirely a tax on income. This is a simplification since in reality taxes are collected in many forms from many different agents. For example, governments collect profit taxes from firms and financial institutions, sales and property taxes from households, and tariffs on traded goods (not included yet). All of these taxes are assumed to be lumped together in the T flow and withdrawn directly from national income.

Tax revenues (TR) can be spent in two separate ways. The TR flow represents transfer payments injected into the household income stream. Transfer payments

include social security paid to retired workers, Medicaid and welfare payments, unemployment, and so on. These are government expenditures that do not exchange for a particular good or service. The second type of expenditure is G . G represents spending by government for the purchase of goods and services produced by firms. It includes defense spending, education, police and fire protection, and so on.

The final monetary flow, shown flowing out of the government, is labeled S_G and refers to government saving. It should be obvious that the money collected by government in the form of taxes need not always equal government expenditures. In the event that tax revenues exceed expenditures, the government would have extra money left over. We imagine that this money would be saved in the financial sector since it is always better to collect interest when possible. Hence we draw the flow of excess funds, government saving (S_G), flowing from government into the financial sector.

We can now represent the flow of funds in and out of the government sector with the following identity:

$$S_G = T - TR - G.$$

When T exceeds the sum of TR and G , the government has extra saving that flows into the financial sector. These funds would then be available to lend out and finance additional investment.

Of course, what is more typical of many governments is for the sum of TR and G to exceed tax revenue, T . In this case, the flow of government saving (S_G) would be negative and would be represented in the diagram as a flow in the opposite direction. In this case, the government would be borrowing money from the financial sector to finance its excess expenditures. We would also say that the government is running a budget deficit.

In short, negative government saving, that is, $S_G < 0$, implies a government budget deficit, which the government finances by borrowing from the financial sector.

Otherwise, positive government saving, that is, $S_G > 0$, implies a government budget surplus, which results either in additions to saving or a repayment of previous debt.

Next, in this version of the circular flow, we can represent the national income identity as the flow of money into firms. In this case, GNP equals the sum of C , I , and

G. This version would only be Accurate when there is no trade with the rest of the world.

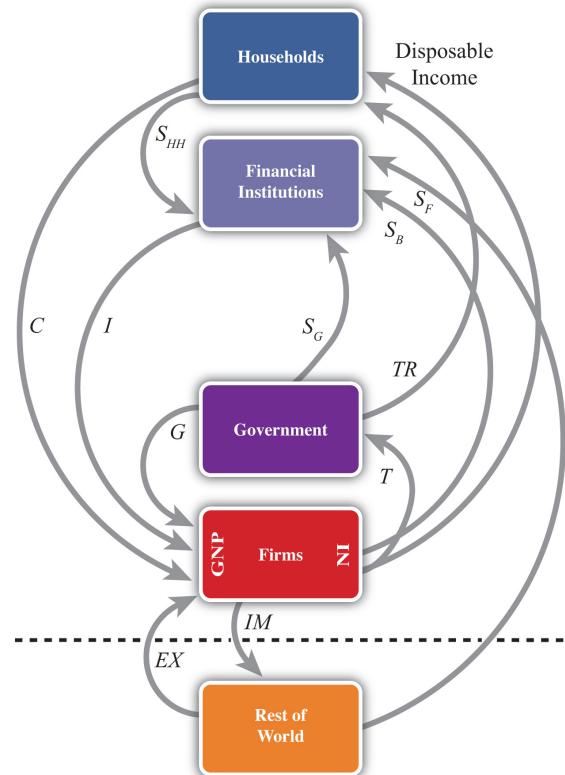
Lastly, with government included, we must rewrite the relationship representing the flows in and out of the financial sector. This now becomes

$$S_{HH} + S_B + S_G = I.$$

This identity says that the sum of household, business, and government saving must equal private investment expenditures.

Circular Flow: Version 4

Figure 13.5 The Circular Flow Adding the RoW



The final circular flow diagram shown in [Figure 13.5 "The Circular Flow Adding the RoW"](#) extends the previous version to include trade flows with the rest of the world. The rest of the world (RoW) is shown at the very bottom of the adjoining diagram, below the dotted line, which represents the border. Trade with the RoW consists first of exports of goods, services, income and transfers, and expenditures on

exports (EX), represented by a flow into firms since money is being used by foreigners to purchase the exported products. Second, imports of goods, services, income and transfers, and imports (IM) are subtracted from firms, resulting in an arrow from firms to the RoW. This adjustment accounts for the fact that measured expenditures made by households, the government, and firms in an open economy will consist of purchases of both domestic and imported goods. Thus the C , I , and G flows will include their purchases of imports, and these should not be included as part of GNP. In essence, the money used to buy imported products is redirected to the foreign firms, hence we have the outflow of money. (For a more complete explanation see [Chapter 13 "National Income and the Balance of Payments Accounts"](#), [Section 13.1 "National Income and Product Accounts"](#).)

This completes the national income identity with all major sectors included and now becomes

$$GNP = C + I + G + EX - IM,$$

which is represented by the flow of money into (and away from) firms on the left side of the diagram. However, as noted elsewhere, $EX - IM$, the balance on the current account, need not be equal to zero. If $EX - IM > 0$, then the country would have a current account (CA) surplus, whereas if $EX - IM < 0$ the country would have a CA deficit.

Consider when $EX - IM < 0$. In this case, more money flows out to purchase imports than flows back in to purchase exports. Essentially, there is a loss of money to the RoW despite some exceptions; however, this money does not remain outside the country. Instead, it is brought right back in and deposited into financial institutions (shown as the S_F flow on the diagram). In other words, it is saved. This saving represents the country's financial account surplus, which is equal and opposite to the CA deficit (see [Chapter 13 "National Income and the Balance of Payments Accounts"](#), [Section 13.5 "Recording Transactions on the Balance of Payments"](#) for a more complete explanation).

The key point is that foreign saving offsets the CA deficit. This can be represented by the relationship showing the inflows and outflows from the RoW, namely,

$$S_F = IM - EX.$$

This says that foreign saving equals the CA deficit. From the perspective of the foreigners, we would refer to S_F as money saved or lent to the domestic country. From the perspective of the domestic country, S_F would be considered money borrowed from the RoW.

Clearly, since a country may run a surplus on trade (i.e., $EX - IM > 0$), S_F could also be negative. In this case, the RoW would either be dissaving, meaning it is withdrawing previously accumulated saving from the domestic country, or the RoW would be borrowing money from the domestic country. This would occur if a domestic bank makes a loan to someone abroad. Alternatively, from the perspective of the domestic country, we can say it is lending money to the RoW when $S_F < 0$.

Finally, the Twin-Deficit Identity

The twin-deficit identity is derived by accounting for the monetary flows in and out of the financial sector in version four of the circular flow. This results in the following identity:

$$S_{HH} + S_B + S_G + S_F = I.$$

This says that the sum of household saving, business saving, government saving, and foreign saving must equal private investment spending. An equivalent version can be written by recalling that household plus business saving equals private saving to get

$$S_P + S_G + S_F = I.$$

The identity is best interpreted by noting that there are four key sources for funds in the financial sector that are not part of the consumption stream. The pool of funds to finance investment can be drawn from households, businesses, the government, or from the RoW. Also, the sum of all funds not used for consumption must be equal to the amount spent on investment goods.

It is important to note that this relationship is an accounting identity. This means that the relationship must be true as long as all variables are measured properly. This is *not* an economic theory, which is a proposition that may or may not be true. In practice, this identity rarely adds up, however, because the variables are not typically measured accurately.

To turn this identity into the “twin-deficit” identity, we must merely take note of several previous definitions. Recall that

$$S_G = T - TR - G,$$

$$S_F = IM - EX,$$

and

$$S_P = S_{HH} + S_B.$$

Plugging these into identity 1 above yields

$$S_P + T - TR - G + IM - EX = I.$$

Reorder these to get the following twin-deficit identity:

$$(S_P - I) + (IM - EX) = (G + TR - T).$$

This is a popular way of writing the twin-deficit identity since it explicitly indicates two deficits. If the second expression $(IM - EX) > 0$, then the country has a current account deficit (i.e., a trade deficit). If the right-hand-side expression $(G + TR - T) > 0$, then the country has a government budget deficit. The expression in total, then, demonstrates that these two deficits are related to each other according to this accounting identity. Indeed, the difference between the government budget deficit and the trade deficit must equal the difference between private saving and investment as shown here:

$$(S_P - I) = (G + TR - T) - (IM - EX).$$

The Twin-Deficit Relationship in the United States and China

Perhaps the best way to get a feel for the twin-deficit relationship in a country is to look at the numbers. [Table 13.7 "U.S. Twin-Deficit Figures \(GDP\), 1997–2008"](#) and [Table 13.8 "China Twin-Deficit Figures \(GDP\), 1997–2007"](#) show values for the twin-deficit identity in the United States and in China over the past ten years or so. All values are presented as a percentage of GDP. Also, because the data on the balance of payments never add up, which results in a statistical discrepancy term, the twin-deficit identity numbers do not add up. To avoid that problem, the private saving numbers presented are not the actual reported values but the values saving would have to be to assure the twin-deficit identity adds up—that is, it is derived as a residual value.

Table 13.7 U.S. Twin-Deficit Figures (GDP), 1997–2008

	$(S_p - I) + \text{Current Account Deficit} = \text{Govt. Budget Deficit}$			
Year	Private Saving* (%)	Investment (%)	Current Account Deficit (%)	Govt. Budget Deficit (%)
2008	13.5	14.0	4.7	4.2
2007	11.7	15.4	5.3	1.6
2006	12.1	16.7	6.1	1.5
2005	12.9	16.5	6.1	2.5
2004	14.0	16.1	5.5	3.4
2003	14.0	15.2	4.8	3.6
2002	13.4	15.1	4.4	2.7
2001	11.6	15.9	3.8	-0.5
2000	11.0	17.7	4.2	-2.4
1999	12.6	17.5	3.2	-1.7
1998	13.8	17.3	2.4	-1.0
1997	15.2	16.7	1.7	0.2

*** Private saving is calculated as a residual.**

Source: U.S. Bureau of Economic Analysis, National Economic Accounts, Frequently Requested NIPA Tables. See U.S. BEA interactive tables for the years indicated at <http://www.bea.gov/national/nipaweb>SelectTable.asp?Popular=Y>.

Table 13.8 China Twin-Deficit Figures (GDP), 1997–2007

	$(S_p - I) + \text{Current Account Deficit} = \text{Govt. Budget Deficit}$			
Year	Private Saving* (%)	Investment (%)	Current Account Deficit (%)	Govt. Budget Deficit (%)
2007	53.0	42.3	-11.3	-0.6
2006	52.8	42.6	-9.4	0.8
2005	51.1	42.7	-7.2	1.2
2004	48.1	43.2	-3.6	1.3

*** Private saving is calculated as a residual.**

Year	$(S_p - I) + \text{Current Account Deficit} = \text{Govt. Budget Deficit}$			
	Private Saving* (%)	Investment (%)	Current Account Deficit (%)	Govt. Budget Deficit (%)
2003	46.0	41.0	-2.8	2.2
2002	43.0	37.9	-2.4	2.6
2001	40.1	36.5	-1.3	2.3
2000	39.5	35.3	-1.7	2.5
1999	39.6	36.2	-1.4	1.9
1998	40.2	36.2	-2.9	1.1
1997	40.6	36.7	-3.1	0.7
* Private saving is calculated as a residual.				

Source: China Data Online, China Statistical Yearbook. See China Statistical Yearbooks located at <http://chinadataonline.org/member/yearbooksp/default.asp?StartYear=1981&EndYear=2009&IFFirst=yes&page=2>.

The twin-deficit numbers reveal some interesting patterns. As of the most recent data (2008), the United States has twin deficits, with a CA deficit of 4.7 percent of GDP and a government budget deficit of 4.2 percent. Since these numbers are almost equal, it is as if the U.S. government deficit, which must be financed with borrowing, is being financed by borrowed funds from abroad. In the previous year, 2007, government borrowing requirements were much lower, at 1.6 percent, but borrowing from foreigners was higher at 5.3 percent. The extra borrowing allowed the U.S. savings rate to remain much lower than the private investment requirement. We can interpret this year as one in which private investment was mostly financed with borrowings from abroad.

The United States has had twin deficits since 2001, when it finished a four-year run with a trade deficit and a government budget surplus. This demonstrates that twin “deficits” do not always arise despite the label used to describe the identity. During the budget surplus years the government was able to retire some of its outstanding debt, but the country also ran CA deficits implying, essentially, borrowings from foreigners. As in 2007, these years also describe periods in which foreign borrowings are used to maintain a higher investment level than can be sustained with the lower national savings rate.

In contrast, consider the twin-deficit numbers calculated in the same way for China during the same period. The differences with the U.S. numbers are striking. The two

things that stand out immediately are the significantly higher values for private saving and investment. Instead of numbers in the midteens in the United States, China's percentages are in the midforties to low fifties. Again, the savings terms are calculated as residuals, so there may be some error there, but nonetheless it is clear that China both saves and invests about three times more than the United States as a percentage of GDP. Because it invests so much more, the implication from the national income identity is that China consumes much less than the United States as a percentage. Indeed, China's consumption figures (not shown) are usually less than 50 percent of GDP.

Indeed, this is why China and many other Asian economies are described as high-saving and low-consuming countries. The United States in comparison is described as a high-consumption country and low-saving country.

The negative number on China's CA deficit in all the years means that China has run a trade surplus. A surplus means it is lending money abroad and forgoing consumption, by another 11 percent in 2007. (This will be explained in more detail in [Chapter 14 "The Whole Truth about Trade Imbalances"](#).) Also, the negative number for China's budget balance means that it was running a government budget surplus in 2007. So in 2007, China had twin surpluses—a much rarer occurrence—rather than twin deficits. In previous years China didn't have twin anything: running trade surpluses that were increasing through the past decade, and government budget deficits.

It is worth reflecting briefly on the large investment and trade surpluses in China in comparison with the United States. The U.S. per capita GDP is about \$47,000. Comprising that per person production is about 15 percent that goes into investment. That still leaves a considerable percentage left for the consumption and government spending that enhance Americans' standard of living. In contrast, China's per capita GDP, in purchasing power parity (PPP) terms, is about \$6,000. Per person, it produces much less than in the United States. But curiously, despite being a much poorer country, the high investment rate means that it consumes and spends on government programs a much smaller percentage of its income than the United States; perhaps as little as \$3,000 per person.

This seems to fly in the face of simple logic. One might expect that a richer country like the United States would save more and consume less since it can do so while still maintaining a high standard of living. For a poorer country like China, we might expect it would save less and try to consume a larger proportion of its income in order to catch up (i.e., in terms of its standard of living) with the rest of the world. Instead, it is the exact opposite.

KEY TAKEAWAYS

- Twin deficits occur when a country has both a current account deficit and a government budget deficit at the same time.
- When twin deficits occur, the sum of net private saving ($S_p - I$) and the current account deficit must equal the government budget deficit.
- A government budget deficit represents a use of funds drawn from the financial sector.
- A trade deficit represents a source of funds for the financial sector.
- Private saving represents a source of funds for the financial sector.
- Private investment represents a use of funds drawn from the financial sector.
- The United States has run twin deficits for the past seven years. It can be reasonably described as a low-investment, low-saving, and high-consumption country.
- China has mostly run trade surpluses and budget deficits in the past decade. It can be reasonably described as a high-investment, high-saving, and low-consumption country.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. An excess of government receipts over expenditures.
 - b. National income minus taxes plus transfer payments.
 - c. The level of government spending when the government deficit is \$100 billion, transfer payments are \$800 billion, and tax revenues are \$1,300 billion.
 - d. The four different sources of saving described in this chapter.
 - e. Of *deficit, surplus, or balance*, the balance on the current account if the expression $IM - EX$ in the twin-deficit identity is positive.
 - f. Of *deficit, surplus, or balance*, the balance on the government budget if the expression $(G + TR - T)$ in the twin-deficit identity is positive.
2. What is the government’s budget balance if government spending is \$40 billion, private saving is \$60 billion, government transfer payments are \$10 billion, private investment is \$80 billion, and tax revenues are \$50 billion? Show your work.
3. Below are the economic data for the fictional country of Sandia. Write out the twin-deficit identity. Verify whether Sandia’s data satisfy the identity.

TABLE 13.9 SANDIA’S ECONOMIC DATA
(BILLIONS OF DOLLARS)

Gross Domestic Product	400
Imports of Goods and Services	140
Investment Spending	20
Private Saving	30
Exports of Goods and Services	100

Government Transfers	40
Government Tax Revenues	140
Government Spending	140
Consumption Spending	280

4. Japan once argued that the main reason the United States had large trade deficits during the 1980s and 1990s was because of its large federal government budget deficit. If the United States wanted to reduce its trade deficit, Japan said, then the United States should reduce its budget deficit. Use the twin-deficit identity to answer the following questions:
- a. Explain what also would have to hold for there to be a direct relationship between budget deficit changes and trade deficit changes.
 - b. Is it possible to account for a reduction in the federal government budget deficit and a simultaneous increase in the current account deficit? Explain.
 - c. Is it possible to reduce the federal government budget deficit, maintain the same level of net private saving (i.e., $S_p - I$), and still experience an increase in the current account deficit? Explain.
5. Explain whether the following economic changes are consistent with the twin-deficit identity. Assume *ceteris paribus*, meaning all other variables in the identity remain fixed.
- a. A \$10 billion increase in the government budget deficit and a \$10 billion increase in the current account deficit.
 - b. A \$50 billion decrease in the government budget deficit and a \$50 billion increase in private investment.
 - c. A \$10 billion increase each in the government budget surplus, the current account deficit, private saving, and private investment.
 - d. A \$30 billion increase in the current account surplus and a \$30 billion increase in the government budget deficit.
6. Refer to the table below to answer the following questions:

- a. Use the twin-deficit identity to fill in the blank values in the table below for the three fictitious countries.

	Private Saving (S_p)	Investment	Current Account Deficit	Government Budget Deficit
Metis	500	500		200
Thebe		150	0	300
Leda	75	100	0	

- b. Which country is best described as financing its government budget deficit with domestic saving?
 c. Which country is best described as financing its government budget deficit with foreign saving?
 d. Which country is best described as financing extra domestic investment with government saving?

13.8 International Investment Position

LEARNING OBJECTIVES

1. Learn how to define and interpret a country's international investment position.
2. Understand how the international investment position is updated from year to year.

A country's **international investment position (IIP)**¹⁰ is like a balance sheet in that it shows the total holdings of foreign assets by domestic residents and the total holdings of domestic assets by foreign residents at a point in time. In the International Monetary Fund's (IMF) financial statistics, these are listed as domestic assets (foreign assets held by domestic residents) and domestic liabilities (domestic assets owned by foreign residents). The financial account balance, whose counterpart is the current account balance, is more like an income statement that shows the changes in asset holdings during the past year. In other words, the financial account balance consists of flow variables since it records changes in the country's asset holdings during the year, while the international asset position of a country consists of stock variables since it records the total value of assets at a point in time.

A country's net international asset position may be in surplus, deficit, or balance. If in surplus, then the value of foreign assets (debt and equity) held by domestic residents exceeds the value of domestic assets held by foreigners. Alternatively, we could say that domestic assets exceed domestic liabilities. This country would then be referred to as a **creditor country**¹¹. If the reverse is true, so that domestic liabilities to foreigners exceed domestic assets, then the country would be called a **debtor country**¹².

10. A measure of the difference between the total value of domestic holdings of foreign assets and the value of foreign assets held in the domestic country.

11. A country whose total domestic assets held abroad exceed total domestic liabilities held by foreigners.

12. A country whose total domestic liabilities held by foreigners exceeds total domestic assets held abroad.

Asset holdings may consist of either debt obligations or equity claims. Debt consists of IOUs (i.e., I owe you) in which two parties sign a contract agreeing to an initial transfer of money from the lender to the borrower followed by a repayment according to an agreed schedule. The debt contract establishes an obligation for the borrower to repay principal and interest in the future. Equity claims represent ownership shares in potentially productive assets. Equity holdings do not establish obligations between parties, at least not in the form of guaranteed repayments. Once ownership in an asset is transferred from seller to buyer, all advantages and disadvantages of the asset are transferred as well.

Debt and equity obligations always pose several risks. The first risk with debt obligations is the risk of possible default (either total or partial). To the lender, default risk means that the IOU will not be repaid at all, that it will be repaid only in part, or that it is repaid over a much longer period of time than originally contracted. The risk of default to the borrower is that future borrowing will likely become unavailable. The advantage of default to the borrower, of course, is that not all the borrowed money is repaid. The second risk posed by debt is that the real value of the repayments may be different than expected. This can arise because of unexpected inflation or unexpected currency changes. Consider inflation first. If inflation is higher than expected, then the real value of debt repayment (if the nominal interest rate is fixed) will be lower than originally expected. This will be an advantage to the borrower, who repays less in real terms, and a disadvantage to the lender, who receives less in real terms. If inflation turns out to be less than expected, then the advantages are reversed. Next, consider currency fluctuations. Suppose a domestic resident, who receives income in the domestic currency, borrows foreign currency in the international market. If the domestic currency depreciates, then the value of the repayments in domestic currency terms will rise even though the foreign currency repayment value remains the same. Thus currency depreciations can be harmful to borrowers of foreign currency. A similar problem can arise for a lender. Suppose a domestic resident purchases foreign currency and then lends it to a foreign resident (note that this is the equivalent of saving money abroad). If the domestic currency appreciates, then foreign savings, once cashed in, will purchase fewer domestic goods and the lender will lose.

The risk of equity purchases arises whenever the asset's rate of return is less than expected. This can happen for a number of different reasons. First, if the equity purchases are direct investment in a business, then the return on that investment will depend on how well the business performs. If the market is vibrant and management is good, then the investment will be profitable. Otherwise, the rate of return on the investment could be negative. All the risk, however, is borne by the investor. The same holds true for stock purchases. Returns on stocks may be positive or negative, but it is the purchaser who bears full responsibility for the return on the investment. Equity purchases can suffer from exchange rate risk as well. When foreign equities are purchased, their rate of return in terms of domestic currency will depend on the currency value. If the foreign currency in which assets are denominated falls substantially in value, then the value of those assets falls along with it.

The U.S. International Investment Position

The United States is the largest debtor nation in the world. This means that its international investment position is in deficit and the monetary value of that deficit is larger than that of any other country in the world. The data for the U.S.

international investment position in 2008 are available in this U.S. BEA international investment position spreadsheet. The data for the U.S. international investment position are available from the Bureau of Economic Analysis, International Economic Accounts, International Investment Position, at http://www.bea.gov/international/xls/intinv08_t1.xls. At market values the preliminary estimate for 2008 is that the U.S. was in debt to the rest of the world in the amount of \$3.469 trillion. (Refer to cell I22 in spreadsheet.) Excluding financial derivatives that refer to interest rate and foreign exchange contracts, the United States was in debt in the amount -\$3.628 trillion (cell I24).

Note that this valuation is the U.S. “net” investment position, meaning that it is the difference between the sum total value of foreign assets owned by U.S. residents (U.S. assets abroad) minus U.S. assets owned by foreigners (foreign-owned assets in the United States). The first of these, U.S. assets abroad, represents our purchases of foreign equities and money we have lent to foreigners. The total value stood at \$19.888 trillion in 2008 using market value methods (cell I26). The second, foreign-owned assets in the United States, represents foreign purchases of U.S. equities and money foreigners have lent to us or, equivalently, that we have borrowed. The total in 2008 stood at \$23.357 trillion (cell I50).

The size of the U.S. debt position causes worry for some. Thirty years ago the United States had a sizable creditor position. However, as a result of trade deficits run throughout the 1980s and 1990s, the United States quickly turned from a net creditor to a net debtor. The changeover occurred in 1989. In the early 1990s, the size of this debt position was not too large compared to the size of the economy; however, by the late 1990s and early 2000s, the debt ballooned. In 2008, the U.S. debt position stood at 24.6 percent of GDP, which interestingly is down slightly from 24.9 percent of GDP in 2002 despite annual current account deficits since then. The reason for these changes is changes in the valuations of assets, as reflected in stock market prices, real estate price changes, and changes in the exchange rate.

Notice in the 2008 BEA IIP spreadsheet that the investment position is derived from the 2007 position in the following way. First, the current account deficit caused an addition to U.S. external debt of \$505 billion (cell D22). Changes in asset prices both here and abroad further increased U.S. external debt by \$720 billion (cell E22). This could be because either real estate prices abroad fell by more than in the United States or security prices abroad fell by more than in the United States. Next, there was another increase of \$583 billion in external U.S. debt because of changes in exchange rates. In this case, an appreciation of the U.S. dollar increased the values of foreign-held U.S. assets and reduced the value of U.S.-held foreign assets. Finally, U.S. external debt decreased by \$479 billion due to other factors that don’t neatly fit into the first two categories. (See footnote 2 in the BEA IIP spreadsheet.)

For several reasons, the debt is not a cause for great worry, although it is growing quickly. First, despite its large numerical size, the U.S. international debt position is still less than 25 percent of its annual GDP. Although this is large enough to be worrisome, especially with a trend toward a future increase, it is not nearly as large as some other countries have experienced in the past. In Argentina and Brazil, international debt positions exceeded 60 percent of their GDPs. For some less-developed countries, international debt at times has exceeded 100 percent of their annual GDP.

A second important point is that much of our international obligations are denominated in our own home currency. This means that when international debts (principal + interest) are paid to foreigners, they will be paid in U.S. currency rather than foreign currency. This relieves the U.S. from the requirement to sell products abroad to acquire sufficient foreign currency to repay its debts. Many other countries that have experienced international debt crises have had great problems financing interest and principal repayments especially when bad economic times make it difficult to maintain foreign sales.

Finally, it is worth noting that, despite the name applied to it, our international “debt” position does not correspond entirely to “debt” in the term’s common usage. Recall that debt commonly refers to obligations that must be repaid with interest in the future. Although a sizable share of our outstanding obligations is in the form of debt, another component is in equities. That means some of the money “owed” to foreigners is simply the value of their shares of stock in U.S. companies. These equities either will make money or will not be based on the success of the business, but they do not require a formal obligation for repayment in the future.

KEY TAKEAWAYS

- The IIP measures the difference between the total value of domestic holdings of foreign assets and the value of foreign assets held in the domestic country. If the IIP is negative, we say the country is a debtor country. If the IIP is positive, we say the country is a creditor country.
- Asset holdings include both debt and equities. Debt involves an obligation to repay principal and interest, whereas equities involve either profit or loss to the foreign asset holder.
- The U.S. IIP stands at \$3.5 trillion in 2008, making the United States the largest debtor nation in the world.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. A complete record of a country’s holdings of foreign assets and foreigners’ holdings of domestic assets at a point in time.
 - b. The term describing a country whose total domestic assets held abroad exceed total domestic liabilities held by foreigners.
 - c. The term describing a country whose total domestic liabilities held by foreigners exceed total domestic assets held abroad.
 - d. The name for the type of asset that establishes an obligation for the borrower to repay principal and interest in the future.
 - e. The name for the type of asset that represents ownership shares in potentially productive assets.

Chapter 14

The Whole Truth about Trade Imbalances

One of the most misinterpreted and misunderstood concepts in international finance is the implication of a country's trade deficit or surplus. Often it is incorrectly presumed that a trade deficit is problematic while a trade surplus is a sign of economic strength. This chapter walks the reader through a thorough investigation of trade imbalances—what they mean and how to interpret them. The chapter concludes that trade deficits can indeed be a big problem for a country, but not always. Trade surpluses can also be a sign of strength, but again, not always. Whether a trade imbalance for a particular country should be viewed as good, bad, or benign depends on many other economic circumstances. This chapter spells out what those circumstances are.

14.1 Overview of Trade Imbalances

LEARNING OBJECTIVE

1. Recognize that trade deficits are not inherently bad and trade surpluses are not inherently good for a country.

There is a popular and pervasive myth about international trade. The myth, simply stated, is that trade deficits are bad and trade surpluses are good. Good or bad for whom, one might ask? Well, for the entire country.

The presence of a trade deficit, or an increase in the trade deficit in a previous month or quarter, is commonly reported as a sign of distress. Similarly, a decrease in a trade deficit, or the presence of or increase in a trade surplus, is commonly viewed as a sign of strength in an economy.

Unfortunately, these perceptions and beliefs are somewhat misguided. In general, it is simply not true that a trade deficit is a sign of a weak economy and a trade surplus is a sign of a strong economy. Merely knowing that a country has a trade deficit, or that a trade deficit is rising, is not enough information to say anything about the current or future prospects for a country—and yet that is precisely how the statistics are often reported.

The truth about trade deficits is that sometimes they are good, sometimes they are bad, but most times, they are benign (i.e., they just don't matter). There are situations in which trade deficits could be interpreted as a sign of a strong thriving economy. There are other situations in which trade deficits could be indicative of economic problems. In most situations, however, trade deficits are not large enough to warrant a positive or negative interpretation. In this case, they should be viewed without interest. These same points apply to trade surpluses as well.

The purpose of this chapter is to explain, clearly and intuitively, the circumstances in which trade imbalances should be interpreted as good and the circumstances in which they are bad. The section will show situations in which trade deficits can indeed lead to long-term harm for an economy. However, it will also show cases in which trade deficits significantly improve a country's long-term economic prospects. We will highlight cases in which trade surpluses are appropriate and a sign of strength for a country, and we will show other cases in which trade

surpluses may correspond to current demise or even an eventual collapse of an economy.

Most important, one should realize after reading this chapter that merely knowing that a country has a trade deficit or surplus is not enough information to say anything substantive about the strength of a country or its economic prospects.

KEY TAKEAWAY

- Trade deficits or trade surpluses can be good, bad, or benign depending on the underlying economic circumstances.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *good, bad, or benign*, this is what the common myth is about the nature of trade deficits.
 - b. Of *good, bad, or benign*, this is what the common myth is about the nature of trade surpluses.
 - c. Of *good, bad, benign, or all of the above*, in general, this is what trade deficits can be.
 - d. Of *good, bad, benign, or all of the above*, in general, this is what trade surpluses can be.
 - e. Of *good, bad, benign, or all of the above*, perhaps most of the time, this is what trade deficits are.
 - f. Of *good, bad, benign, or all of the above*, perhaps most of the time, this is what trade surpluses are.

14.2 Trade Imbalances and Jobs

LEARNING OBJECTIVE

1. Learn why trade deficits may not be related to job losses in a country.

One of the main reasons trade deficits are considered deleterious is because of a common argument that trade deficits result in job losses. The rationale behind this argument is simple and convincing. There are two parts to the story that begin with the definition of a trade deficit.

First, a trade deficit arises whenever imports exceed exports. One simple reason for an imbalance of this kind is that imports are too large or at least larger than they would be under balanced trade. The most common reason offered in developed countries for why imports are too large is that low import prices arise because less-developed countries have exceedingly low wages paid to workers, lax health and safety standards, or more lenient environmental policies, all of which contribute to a veritable flood of imports.

The effect of excessive imports is said to be the purchase of cheaper foreign goods by domestic consumers rather than purchasing the slightly more expensive domestic varieties. As demand for domestic firms' products falls, these firms are forced to downsize, resulting in the layoff of domestic workers. Thus it is said that trade deficits cause the loss of domestic jobs.

The second story argues that the reason imports exceed exports is because exports are too low; they are smaller than they should be. The most common reason given for low exports, especially in the developed countries, is the relatively high barriers to trade in developing countries. Although many countries participate in the World Trade Organization (WTO), the average applied tariffs still remain considerably higher in developing countries.

The effect of insufficient exports is that products that could be produced and sold abroad are not produced and sold abroad because of the barriers to trade. If the barriers were only removed, then exports would expand and jobs would be created in the country.

Thus since both of these stories can operate simultaneously, most observers are convinced that trade deficits indeed will cause job losses. Turn the deficit around, perhaps so much so as to induce a trade surplus, and this logic suggests that more jobs will be created.

This argument is very convincing because there is an element of truth to it. Changes in import and export patterns will certainly have competitive impacts on some industries and could produce temporary job losses. However, this doesn't mean that a country with a trade deficit generates fewer overall jobs than a country with a trade surplus. Nor does it mean that increases in a country's trade deficit will necessarily lead to economy-wide job losses.

One reason job losses may not occur has to do with the deceptive nature of the previous job loss stories. The stories are convincing as far as they go, but unfortunately, they don't go far enough. In other words, the job loss stories have some validity, but they are incomplete; they don't tell the full story, and as a result they tend to mislead.

The rest of the story (as Paul Harvey would have said) is to recognize that when trade deficits arise on the current account, there is an equal and opposite trade surplus on the financial account of the balance of payments. A financial account surplus means that foreigners are purchasing domestic assets. Some of these purchases consist of equities such as stocks and real estate, while other asset purchases involve the lending of money as when foreigners purchase a government bond. In any case, that money flows back into the deficit country and ultimately is spent by someone. That someone could be the previous holder of the real estate or it could be the domestic government. When it is spent, it creates demands for goods and services that in turn create jobs in those industries.

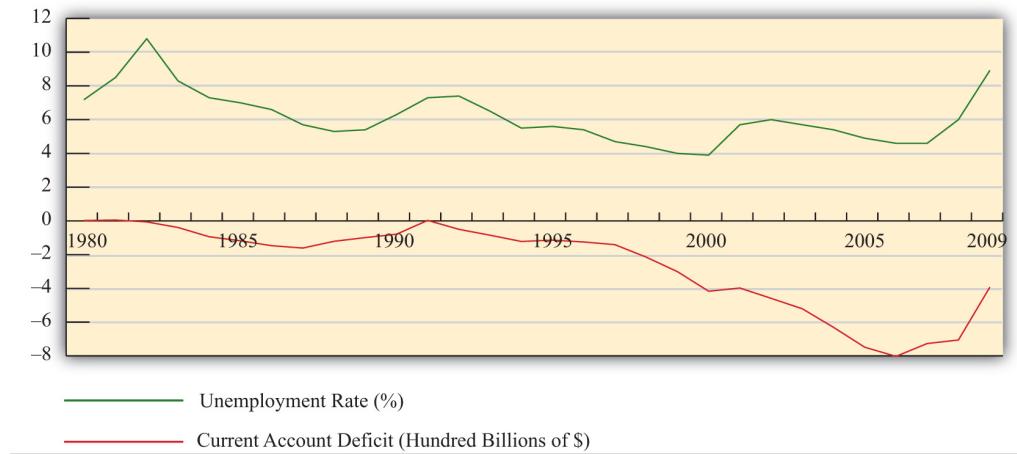
Now consider for a moment the following thought experiment. Suppose we could instantly change the behavior of the foreign lenders generating the financial account surplus (and the related trade deficit). Suppose they decide at once *not* to lend the money to the government or *not* to purchase real estate but instead decide to purchase domestic goods. The increase in goods purchases by foreigners would imply that export demand and hence exports will rise. Indeed, they will rise sufficiently to eliminate the trade deficit. And because of the increase in exports, jobs will be created in the export industries. However, at the same time export jobs are created, other jobs in the economy are being lost. That's because now less money is there to purchase the real estate or to lend to the government. Thus the elimination of the trade deficit doesn't create jobs in the aggregate, but it will change which sectors have more and less demand for its products. In other words,

changes in the trade deficit will ultimately affect only where the jobs are in the economy (i.e., in which industries), not how many jobs there are.

The one exception to this, and one of the main reasons the job loss stories remain so convincing, is when there are rapid changes in the trade deficit or surplus. Rapid changes, like the thought experiment above, would require adjustments of workers between industries. During that adjustment process, some workers will be temporarily unemployed. If that adjustment involves an increase in the trade deficit or a decrease in the trade surplus, the temporary jobs effect will be very noticeable in the tradable products industries. However, if the adjustment involves a decrease in the deficit or an increase in the surplus, then the job losses will more likely occur in the nontradable products sectors and it will be difficult to connect those job losses to the changes in the trade balance.

To provide some validation of this point—that is, that changes in the trade balances do not have effects on the aggregate number of jobs in an economy—consider [Figure 14.1 "U.S. Trade Deficits and Unemployment, 1980–2009"](#), showing two U.S. macroeconomic variables plotted over the past twenty years: the current account balance and the national unemployment rate. Now if the jobs stories suggesting that trade deficits cause job losses were true, we might expect to see an inverse relationship between the trade balance and the unemployment rate. Alternatively, if an increase in a country's trade deficit causes job losses in the economy, we might expect an increase in the unemployment rate to occur as well. Similarly, a decrease in the trade deficit should create jobs and lead to a decrease in the unemployment rate.

Figure 14.1 U.S. Trade Deficits and Unemployment, 1980–2009



Interestingly, what [Figure 14.1 "U.S. Trade Deficits and Unemployment, 1980–2009"](#) shows is that during the periods when the U.S. trade deficit is rising (i.e., the trade balance is falling), the unemployment rate is falling; whereas when the trade deficit is falling, the unemployment rate is rising. This is precisely the opposite effect one would expect if the job-loss stories of trade deficits were true.

Of course this evidence does not prove that trade deficits will reduce unemployment in every country in all circumstances. However, the evidence does suggest that it is inappropriate to jump to the popular conclusion that trade deficits are bad for jobs and thus bad for the economy.

KEY TAKEAWAYS

- Trade deficits are often incorrectly presumed to cause job losses in an economy.
- The job-loss stories suggest that trade deficits arise due to excessive imports or insufficient exports and that by eliminating a deficit a country can create jobs in the economy.
- The job-loss story is incomplete though because it ignores the demand and jobs caused by the financial account surplus.
- When all effects of trade imbalances are accounted for, trade deficits may cause no more than temporary job losses in transition but not affect the aggregate level of jobs in an economy.
- Evidence from the United States over the past twenty years is used to show that the relationship between trade deficits and the unemployment rate is the opposite from what the popular “trade deficits cause job losses” stories would suggest.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *too large, too small, or just right*, concerns about trade deficits sometimes suggest this about imports.
 - b. The import effect on trade deficits is sometimes said to be caused by this wage phenomenon in foreign countries.
 - c. The import effect on trade deficits is sometimes said to be caused by this environmental legal phenomenon in foreign countries.
 - d. Of *too large, too small, or just right*, concerns about trade deficits sometimes suggest this about exports.
 - e. The export effect on trade deficits is sometimes said to be caused by this trade barrier phenomenon in foreign countries.
 - f. The “trade deficits cause job losses” story ignores the effects of international transactions recorded on this balance of payments account.
 - g. Of *increase, decrease, or stay the same*, this has been the typical corresponding change in the U.S. unemployment rate whenever the U.S. trade deficit was rising since 1980.

14.3 The National Welfare Effects of Trade Imbalances

LEARNING OBJECTIVES

1. Understand the long-term implications of trade imbalances.
2. Identify conditions under which trade imbalances are detrimental, beneficial, or benign.

In this section, a series of simple scenarios (or stories) are presented to demonstrate how the well-being of a country may be affected when it runs a trade imbalance. The scenarios compare national output with domestic spending over two periods of time under alternative assumptions about the country's trade imbalance and its economic growth rate between the two periods. After each aggregate scenario is presented, we also provide an analogous situation from the point of view of an individual. Finally we present an evaluation of each scenario and indicate countries that may be displaying similar trade patterns.

Two periods are used as a simple way to introduce the dynamic characteristics of trade imbalances. The amount of time between the two periods can be varied to provide alternative interpretations. Thus the two periods could be labeled as *today and tomorrow*, *this year and next year*, or *this generation and next generation*.

We assume that all trade imbalances correspond to debt obligations or IOUs (i.e., I owe you). In other words, the financial account imbalances that offset the trade imbalances will be interpreted as international borrowing and lending rather than, say, foreign direct investment flows or real estate purchases.

Afterward, we will comment on how the interpretations of these scenarios may change with the alternative type of asset flow.

National welfare is best measured by the amount of goods and services that are “consumed” by households. What we care about, ultimately, is the standard of living obtainable by the average citizen, which is affected not by how much the nation produces but by how much it consumes. Although **gross domestic product (GDP)**¹ is often used as a proxy for national welfare, it is an inadequate indicator for many reasons, especially when a country runs trade imbalances. To quickly see why, consider the extreme situation in which a country runs the largest trade surplus possible. This would arise if a country exports all of its GDP and imports

1. Measures the total value of all goods and services produced by a country during a year.

nothing. The country's trade surplus would then equal its GDP, but the citizens in the country would have no food, clothing, or anything else to consume. The standard of living would be nonexistent.

To avoid this problem we use **domestic spending (DS)**², or the sum of domestic consumption, investment, and government spending, as a proxy for national welfare. More formally, let

$$DS = C + I + G,$$

where C , I , and G are defined as in the national income accounts. Recall from [Chapter 16 "Interest Rate Parity"](#) that C , I , and G each can be segmented into spending on domestically produced goods and services and spending on imported goods and services. Thus domestic spending includes imported goods in the measure of national welfare. This is appropriate since imported goods are consumed by domestic citizens and add to their well-being and standard of living.

One problem with using domestic spending as a proxy for average living standards is the inclusion of investment (note that this problem would also arise using GDP as a proxy). Investment spending measures the value of goods and services used as inputs into the productive process. As such, these items do not directly raise the well-being of citizens, at least not in the present period. To clarify this point, consider an isolated, self-sufficient corn farmer. Each year the farmer harvests corn, using part of it to sustain the family during the year, while allocating some of the kernels to use as seed corn for the following year. Clearly, the more kernels the farmer saves for next year's crop, the less corn the family will have to consume this year. As with the farmer, the same goes for the nation: the more that is invested today, the lower will be today's standard of living, *ceteris paribus*. Thus we must use domestic spending cautiously as a measure of national welfare and take note of changes in investment spending if it occurs.

The analysis below will focus on the interpretation of differences between national income (GDP) and domestic spending under different scenarios concerning the trade imbalance. The relationship between them can be shown by rewriting the national income identity.

The national income identity is written as

$$GDP = C + I + G + EX - IM.$$

Substituting the term for domestic spending yields

2. The sum of domestic consumption, investment, and government spending. Often used as a proxy for national welfare.

$$GDP = DS + EX - IM,$$

and rearranging it gives

$$EX - IM = GDP - DS.$$

The last expression implies that when a country has a current account (or trade) surplus, GDP must exceed domestic spending by the equivalent amount. Similarly, when a country has a trade deficit, domestic spending exceeds GDP.

Note that to be completely accurate, we should use growth national product (GNP) rather than GDP in the analysis. This is because we are interpreting $EX - IM$ as the current account balance that includes income payments and receipts. With income flows included on the trade side, the measure of national output we get is GNP not GDP. Because conceptually both are measures of national output, we will use GNP in everything that follows in this section.

Case 1: No Trade Imbalances; No GNP Growth between Periods

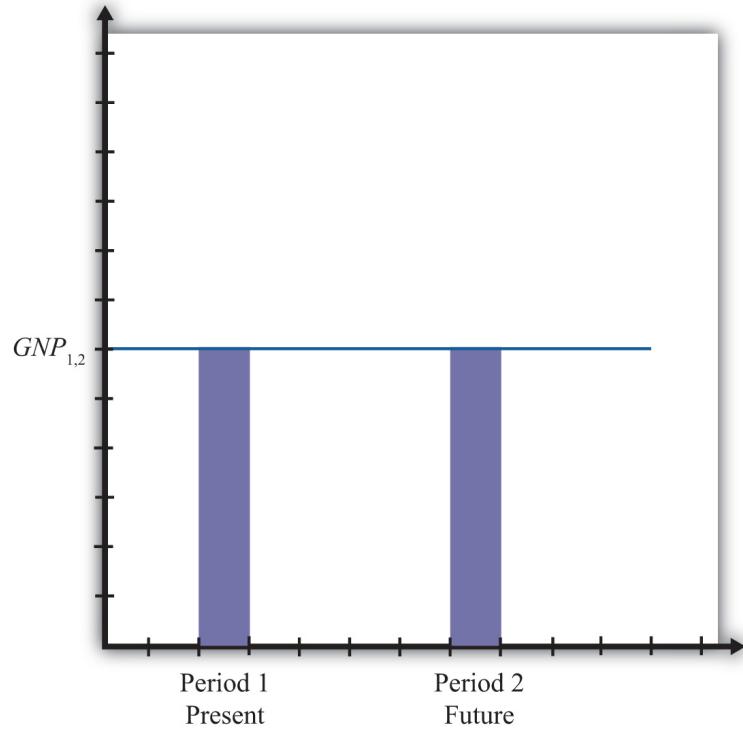
Case 1, what we will call the base case, is used to demonstrate how GNP compares with domestic spending in the simplest scenario. Here we assume that the country does not run a trade deficit or surplus in either of the two periods and that no GNP growth occurs between periods. No trade imbalance implies that no net international borrowing or lending occurs on the financial account. The case mimics how things would look if the country were in autarky and did not trade with the rest of the world.

Note from [Figure 14.2 "Case 1"](#) that domestic spending, shown as the aqua bar graph, is exactly equal to GNP in both periods. Since domestic spending is used to measure national welfare, we see that the average standard of living remains unchanged between the two periods. Overall, nothing very interesting happens in this case, but it will be useful for comparison purposes.

Figure 14.2 Case 1

$$DS = C + I + G,$$

GNP



The Individual Analogy

Consider an individual named Rajiv. For an individual, GNP is analogous to Rajiv's annual income since his income represents the value of goods and services produced with his labor services. Domestic spending is analogous to the value of the goods and services purchased by Rajiv during the year. It corresponds to Rajiv's consumption of goods and services that serves as a proxy for his welfare level. Trade for an individual occurs whenever a transaction occurs with someone outside his household.

Let's assume for simplicity that Rajiv earns \$30,000 per year. The assumption of no GNP growth in the base case implies that he continues to earn \$30,000 in the second period and thus experiences no income growth. The assumption of no trade imbalances implies that Rajiv engages in no borrowing or lending outside of his household. That implies that he spends all of his income on consumption goods and thus purchases \$30,000 worth of goods and services. This level of consumption remains the same in both periods, implying that his standard of living is unchanged.

Another way of interpreting balanced trade for an individual is to imagine that he exports \$30,000 worth of labor services and afterward imports \$30,000 worth of consumption goods and services. Since exports equal imports, trade is balanced.

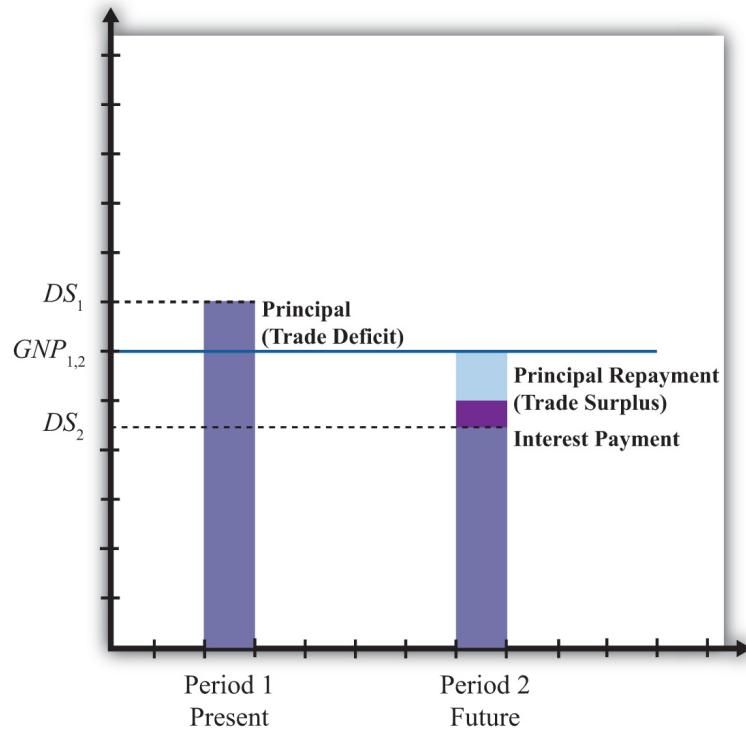
Case 2: Current Account Deficit Period 1; No GDP Growth between Periods

In this case, we assume that the country runs a current account (or trade) deficit in the first period. We'll also assume that the resultant financial account surplus corresponds to borrowing from the rest of the world, rather than asset purchases. These borrowed funds are assumed to be repaid in their entirety in the second period. In other words, we'll assume that loans are taken out in the first period and that the principal and interest are repaid completely in the second period. We also assume that there is no GNP growth between periods.

As shown in [Figure 14.3 "Case 2"](#), the trade deficit in the first period implies that domestic spending, DS_1 , exceeds GNP_1 . The difference between DS_1 and GNP_1 represents the current account deficit as well as the value of the outstanding principal on the foreign loans. The extra consumption the country can enjoy is possible because it borrows funds from abroad and uses them to purchase extra imports. The result is the potential for a higher standard of living in the country in the period in which it runs a current account deficit if the extra funds are not directed into domestic investment.

Figure 14.3 Case 2

$$DS = C + I + G, \\ GNP$$



In the second period, the borrowed funds must be repaid with interest. The repayment reduces domestic spending below the level of GNP by the amount of the principal and interest repayment as shown by the light-colored areas in the diagram. In actuality, the interest repayment component may be included as part of domestic spending since interest represents a payment for services received—those services being the privilege of consuming earlier. However, since this service is unlikely to raise one's standard of living in period two, we have excluded it from domestic spending. Since GNP does not change between the two periods, DS₂ will lie below GNP₁. What this means is that the average standard of living can fall during the period in which the loan repayment is being made.

This outcome highlights perhaps the most important concern about trade deficits. The fear is that large and persistent trade deficits may require a significant fall in living standards when the loans finally come due. If the periods are stretched between two generations, then there is an intergenerational concern. A country running large trade deficits may raise living standards for the current generation, only to reduce them for the next generation. It is then as if the parents' consumption binge is being subsidized by their children.

The Individual Analogy

In case 2, our individual, Rajiv, would again have a \$30,000 income in two successive periods. In the first period, suppose Rajiv borrows money, perhaps by running up charges on his credit card. Suppose these charges amount to \$5,000 and that the interest rate is a generous 10 percent. Assuming Rajiv does not save money in the first period, his consumption level in the first period would be the sum of his income and his borrowed funds. Thus he would enjoy \$35,000 worth of goods and services reflecting a standard of living higher than his actual income.

In the second period, Rajiv must pay back the \$5,000 in loans plus the interest charges, which, at a 10 percent interest rate, would amount to \$500. Thus \$5,500 of Rajiv's \$30,000 income would go toward debt repayment, leaving him with only \$24,500 to spend on consumption.

In this case, extra consumption, or a higher living standard in period one, is achieved by sacrificing a lower living standard in the future.

Note that in the first period Rajiv imports more goods and services in consumption than he exports in terms of labor services. Hence, this corresponds to a trade deficit. In the second period, Rajiv imports fewer goods and services in consumption than the labor services he exports; hence, this corresponds to a trade surplus.

Evaluation

Case two reflects legitimate concerns about countries that run large or persistent trade deficits. The case highlights the fact that trade deficits, which arise from international borrowing, may require a reduced average standard of living for the country in the future when the loans must be repaid.

An example of this situation would be Mexico during the 1970s and 1980s. Mexico ran sizeable current account deficits in the 1970s as it borrowed liberally in international markets.

In the early 1980s, higher interest rates reduced its ability to fulfill its obligations to repay principal and interest on its outstanding loans. Their effective default precipitated the third world debt crisis of the 1980s. During the 1980s, as arrangements were made for an orderly, though incomplete, repayment of Mexico's loans, the country ran sizeable current account surpluses. As in case two here, Mexico's current account deficits in the 1970s allowed it to raise its average living

standards, above what would have been possible otherwise, while its current account surpluses in the 1980s forced a substantial reduction in living standards.

It is worth emphasizing that current account deficits are not detrimental in the periods in which the deficits are occurring. In fact, current account deficits correspond to higher consumption, investment, and government spending levels than would be possible under balanced trade. Instead, current account deficits pose a problem only when the debt repayment occurs, which is when the country is running current account surpluses. Trade deficits raise national welfare in the periods in which they occur, while trade surpluses reduce welfare in the periods in which they occur.

In other words, in terms of the national welfare effects, the problem here isn't large or persistent trade deficits but rather the large and persistent trade surpluses that might arise in the future as a result.

It is also worth noting that trade deficits in this case need not be a problem in the long run if they are not too large. Just as an individual may make a choice to substitute future consumption for present consumption, so might a nation. For example, an individual may reasonably decide while young to take exotic vacations, engage in daredevilish activities, or maybe purchase a fast car, even if it means taking out sizeable loans. Better to enjoy life while healthy, he may reason, even if it means that he will have to forgo similar vacations or activities when he is older. Similarly, a nation, through an aggregation of similar individual decisions, may "choose" to consume above its income today even though it requires reduced consumption tomorrow. As long as the future reduced consumption "costs" are borne by the individuals who choose to overconsume today, deficits for a nation need not be a problem. However, if the decision to overconsume is made through excessive government spending, then the burden of reduced consumption could fall on the future generation of taxpayers, in which case there would be an intergenerational welfare transfer.

Case 3: Current Account Deficit Period 1; Positive GDP Growth between Periods

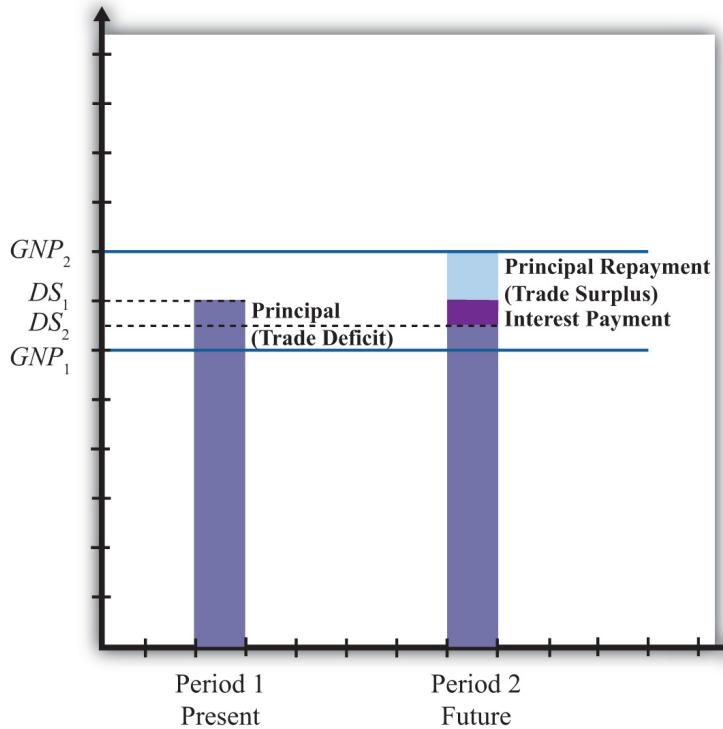
In the third case, we assume, as in case 2, that the country runs a trade deficit in the first period, that the trade deficit corresponds to borrowing from the rest of the world, and that in period two all the loans are repaid with interest. What differs here is that we will assume GNP growth occurs between the first and second periods. As we'll see, growth can significantly affect the long-term effects of trade deficits.

In [Figure 14.4 "Case 3"](#), note that the first period domestic spending (DS_1) lies above GNP in the first period (GNP_1). This arises because a trade deficit implies that the country is borrowing from the rest of the world, allowing it to spend (and consume) more than it produces.

Figure 14.4 Case 3

$$DS = C + I + G,$$

GNP



In the second period, we assume that GNP has grown to GNP_2 as shown in the graph. The principal and interest from first period loans are repaid, which lowers domestic spending to DS_2 . Note that since domestic spending is less than GNP_2 , the country must be running a trade surplus. Also note that the trade surplus implies that consumption and the average standard of living are reduced below the level that is obtainable with balanced trade in that period. In a sense, the trade deficit has a similar long-term detrimental effect as in case two.

However, it is possible that the first period trade deficit, in this case, may actually be generating a long-term benefit. Suppose for a moment that this country's balanced trade outcome over two periods would look like the base case. In that case, balanced trade prevails but no GDP growth occurs, leaving the country with the same standard of living in both periods. Such a country may be able to achieve an

outcome like case three if it borrows money from the rest of the world in period one—thus running a current account deficit—and uses those funds to purchase investment goods, which may in turn stimulate GNP growth. If GNP rises sufficiently, the country will achieve a level of domestic spending that exceeds the level that would have been obtained in the base case.

Indeed, it is even possible for a country's standard of living to be increased in the long term entirely because it runs a trade deficit. In case three, imagine that all the borrowed funds in period one are used for investment. This means that even though domestic spending rises, the average standard of living would remain unchanged relative to the base case because investment goods generate no immediate consumption pleasures. In period two, the higher level of domestic spending may be used for increased consumption that would cause an increase in the country's average living standards. Thus the country is better off in both the short term and long term with the unbalanced trade scenario compared to the balanced trade case.

The Individual Analogy

The third case is analogous to our individual Rajiv with, say, a \$30,000 income in period one. The trade deficit in the first period means that he borrows money using his credit card to purchase an additional, say, \$5,000 worth of “imported” consumption goods. Thus in period one the person's consumption and standard of living are higher than reflected by his income.

In the second period, the GNP rises, corresponding to an increase in Rajiv's income. Let say that his income rises to \$40,000 in the second period. We'll also assume that all credit card loans must be repaid along with 10 percent interest charges in the second period. Consumption spending for Rajiv is now below his income. Subtracting the \$5,000 principal repayment and the \$500 interest payment from his \$40,000 income yields consumption of \$34,500.

The investment story above is similar to the case in which an individual takes out \$5,000 in student loans in period one and earns an advanced degree that allows him to acquire a better-paying job. Assuming the educational investment does not add to his consumption pleasures (a seemingly reasonable assumption for many students), his welfare is unaffected by the additional spending that occurs in period one. However, his welfare is increased in period two since he is able to consume an additional \$4,500 worth of goods and services even after paying back the student loans with interest.

Evaluation

The lesson of case three is that trade deficits, even if large or persistent, will not cause long-term harm to a nation's average standard of living if the country grows rapidly enough. Rapid economic growth is often a cure-all for problems associated with trade deficits.

In some cases, it is possible for growth to be induced by investment spending made possible by borrowing money in international markets. A trade deficit that arises in this circumstance could represent economic salvation for a country rather than a sign of economic weakness.

Consider a less-developed country. Countries are classified as less developed because their average incomes are very low. Indeed, although many less-developed countries, or LDCs, have a small, wealthy upper class, most of the population lives in relative poverty. Individuals who are poor rarely save very much of their incomes, therefore, LDCs generally have relatively small pools of funds at home that can be used to finance domestic investment. If investment is necessary to fuel industrialization and economic growth, as is often the case especially in early stages of development, an LDC might be forced to a slow or nonexistent growth path if it restricts itself to balanced trade and limits its international borrowing.

On the other hand, if an LDC borrows money in international financial markets, it will run a trade deficit by default. If these borrowed funds are used for productive investment, which in turn stimulates sufficient GDP growth, then the country may be able to raise average living standards even after repaying the principal and interest on international loans. Thus trade deficits can be a good thing for less-developed countries.

The same lesson can be applied to the economies in transition in the former Soviet bloc. These countries suffered from a lack of infrastructure and a dilapidated industrial base after the collapse of the Soviet Union. One obvious way to spur economic growth in the transition is to replace the capital stock with new investment: build new factories, install modern equipment, improve the roads, improve telecommunications, and so on. However, with income falling rapidly after the collapse, there were few internal sources to fund this replacement investment. It was also not obvious which sectors were the best to invest in. Nevertheless, one potential option was for these countries to borrow funds on international financial markets. Trade deficits that would occur under this scenario could be justified as an appropriate way to stimulate rapid economic growth.

Of course, just because trade deficits *can* induce economic growth and generate long-term benefits for a country doesn't mean that a trade deficit *will* spur long-term economic growth. Sometimes investments are made in inappropriate industries. Sometimes external shocks cause once profitable industries to collapse. Sometimes borrowed international funds are squandered by government officials and used to purchase large estates and big cars. For many reasons good intentions, and good theory, do not always produce good results. Thus a country that runs large and persistent trade deficits, hoping to produce the favorable outcome shown in case three, might find itself with the unfavorable outcome shown in case two.

Finally, a country running trade deficits could find itself with the favorable outcome even if it doesn't use borrowed international funds to raise domestic investment. The United States, for example, has had rather large trade deficits since 1982. By the late 1980s, the United States achieved the status of the largest debtor nation in the world. During the same period, domestic investment remained relatively low especially in comparison to other developed nations in the world. One may quickly conclude that since investment was not noticeably increased during the period, the United States may be heading for the detrimental outcome. However, the United States maintained steady GNP growth during the 1980s and 1990s, except during the recession year in 1992. As long as growth proceeds rapidly enough, for whatever reason, even a country with persistent deficits can wind up with the beneficial outcome.

Case 4: Current Account Surplus Period 1; No GDP Growth between Periods

In this case, we assume that the country runs a trade surplus in the first period and that no GDP growth occurs between periods. A surplus implies that exports exceed imports of goods and services and that the country has a financial account deficit. We will assume that the financial account deficit corresponds entirely to loans made to the rest of the world. We can also refer to these loans as savings, since the loans imply that someone in the country is forgoing current consumption. In the future, these savings will be redeemed along with the interest collected in the interim. We shall assume that all of these loans are repaid to the country with interest in the second period.

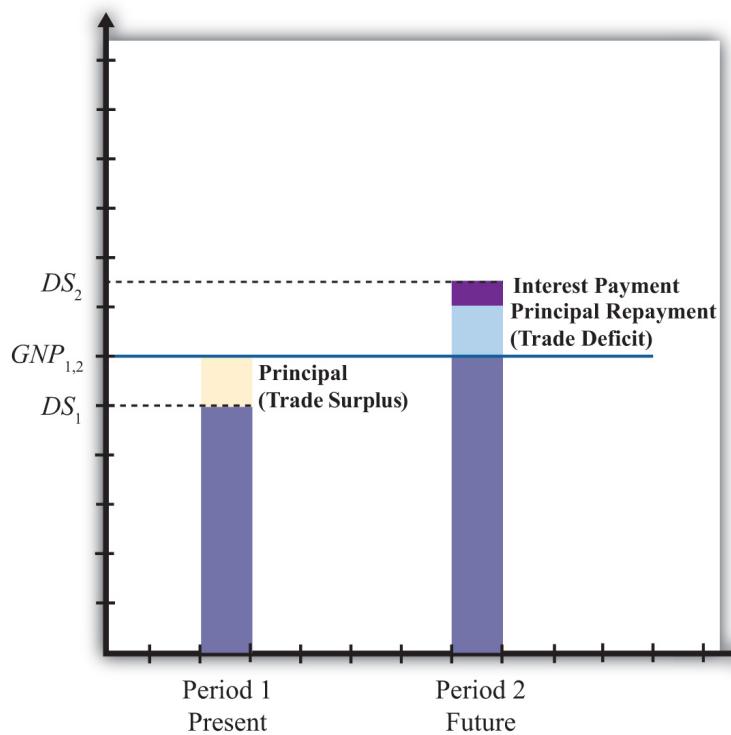
In [Figure 14.5 "Case 4"](#), we see that in the first period, when the trade surplus is run, domestic spending (DS_1) is less than national income or GDP. This occurs because the country is lending rather than consuming some of the money available from production. The excess of exports over imports represents goods that could have been used for domestic consumption, investment, and government spending but are instead being consumed by foreigners. This means that a current account

surplus reduces a country's potential for consumption and investment below what is achievable in balanced trade. If the trade surplus substitutes for domestic consumption and government spending, then the trade surplus will reduce the country's average standard of living. If the trade surplus substitutes for domestic investment, average living standards would not be affected, but the potential for future growth can be reduced. In this sense, trade surpluses can be viewed as a sign of weakness for an economy, especially in the short run during the periods when surpluses are run. Surpluses can reduce living standards and the potential for future growth.

Figure 14.5 Case 4

$$DS = C + I + G,$$

GNP



Nevertheless, this does not mean that countries should not run trade surpluses or that trade surpluses are necessarily detrimental over a longer period. As shown in the diagram, when period two arrives the country redeems its past loans with interest. This will force the country to run a trade deficit, and domestic spending (DS_2) will exceed GDP. The trade deficit implies imports exceed exports, and these additional imports can be used to raise domestic consumption, investment, and government spending. If the deficit leads to greater consumption and government spending, then the country's average standard of living will rise above what is

achievable in balanced trade. If the deficit leads to greater investment, then the country's potential for GDP growth in the third period (not shown) is enhanced.

Briefly, this case describes the situation in which a country forgoes first period consumption and investment so that in period two it can enjoy even greater consumption and investment.

The Individual Analogy

Consider our individual, Rajiv, who has an annual income of \$30,000 over two periods. This corresponds to the constant GDP in the above example. Rajiv would run a trade surplus in period one if he lends money to others. One way to achieve this is simply to put money into a savings account in the local bank. Suppose Rajiv deposits \$5,000 into a savings account. That money is then used by the bank to make loans to other individuals and businesses. Thus in essence Rajiv is making loans to them with the bank acting as an intermediary. The \$5,000 also represents money that Rajiv does not use to buy goods and services. Thus in period one Rajiv exports \$30,000 of labor services, but imports only \$25,000 of consumption goods. The excess is loaned to others so that they may be consumed instead in the first period. It is clear that Rajiv's standard of living at \$25,000 is lower in the first period than the \$30,000 he could have achieved had he not deposited money into savings.

In the second period, we imagine that Rajiv again earns \$30,000 and withdraws all the money plus interest from the savings account. Suppose he had earned 10 percent interest between the periods. In this case, his withdrawal would amount to \$5,500. This means that in period two Rajiv can consume \$35,500 worth of goods and services. This outcome also implies that Rajiv's domestic spending capability exceeds his income and so he must be running a trade deficit. In this case, Rajiv's imports of goods and services at \$35,500 exceed his exports of \$30,000 worth of labor services; thus he has a trade deficit.

Is this outcome good or bad for Rajiv? Most would consider this a good outcome. One might argue that Rajiv has prudently saved some of his income for a later time when he may have a greater need. The story may seem even more prudent if Rajiv suffered a significant drop in income in the second period to, say, \$20,000. In this case, the savings would allow Rajiv to maintain his consumption at nearly the same level in both periods despite the shock to his income stream. This corresponds to the words of wisdom that one should *save for a rainy day*. Savings can certainly allow an individual to smooth his consumption stream over time.

Alternatively, one might consider the two periods of the story to be middle age and retirement. In this case, it would make sense to save money out of one's income in

middle age so that one can draw on those savings and their accumulated earnings during retirement when one's income has fallen to zero.

On the other hand, excessive saving in the first period might make Rajiv seem miserly. Few people would advise that one save so much as to put oneself into poverty or to reduce one's living standard below some reasonable norm. Excessive prudence can seem inappropriate as well.

Evaluation

The prime example of a country that mimics the first period of case four is Japan during the 1980s and 1990s. Japan ran sizeable trade surpluses during those two decades. As this story suggests, the flip side of the trade surplus is a financial account deficit that implied a considerable increase in the amount of loans that Japan made to the rest of the world. Although Japan's trade surplus has often been touted as a sign of strength, an important thing to keep in mind is that Japan's trade surpluses implied lower consumption and government purchases, and thus a lower standard of living than would have been possible with balanced trade. Although trade surpluses can also result in lower investment, this effect was not apparent for Japan. During those two decades of investment, spending as a percentage of GDP always exceeded 25 percent, higher than most other developed countries.

These surpluses may turn out to be especially advantageous for Japan as it progresses in the twenty-first century. First of all, it is clear that Japan's surpluses did not usher in an era of continual and rapid GDP growth. By the early 1990s, Japan's economy had become stagnant and finally began to contract by 1998. However, rather than allowing a decline in GDP to cause a reduction in living standards, Japan could use its sizeable external savings surplus to maintain consumption at the level achieved previously. Of course, this would require that Japan increase its domestic consumption and begin to run a trade deficit, two things that did not occur even by 2009.

In another respect, Japan's trade surpluses may be advantageous over the longer run. Japan, along with most other developed nations, will experience a dramatic demographic shift over the next three decades. Its retired population will continue to grow as a percentage of the total population of the baby boomers reach retirement and people continue to live longer. The size of the Japan's working population will consequently decline as a percentage of the population. This implies an increasing burden on Japan's pay-as-you-go social retirement system as a smaller number of workers will be available per retiree to fund retiree benefits. If at that time Japan draws down its accumulated foreign savings and runs trade deficits, it will be able to boost the average consumption level of its population while

reducing the need to raise tax burdens to fund its social programs. Of course, this outcome may never be realized if Japan's economy does not rebound strongly from its recent stagnant condition.

Overall, regardless of the outcome, Japan's economy today, faced with a potentially severe recession, is certainly in a stronger position by virtue of its accumulated foreign savings than it would be if it had run trade deficits during the past two decades.

Summary

These stories suggest that trade imbalances, when evaluated in terms of their momentary effects and their long-term economic consequences, can be either good, bad, or benign, depending on the circumstances.

Trade deficits may signal excessive borrowing that could in the future lead to possible default, or worse, an excessive reduction in living standards needed to repay the accumulated debt. In this case, the trade deficit is clearly bad for the nation. Alternatively, trade deficits may represent a country that is merely drawing down previously accumulated foreign savings or selling other productive assets, in which case there is no potential for default or reduced living standards in the future. Here, the trade deficit is either immaterial or even beneficial in that the nation is able to achieve a higher current living standard because of the deficit. Trade deficits might also make an expansion of domestic investment possible, which could spur future economic growth sufficiently to make repayment consistent with growing living standards. In this case, trade deficits are clearly good as they stimulate future economic prosperity. Finally, in a free market economy, trade deficits may simply reflect the aggregated choices of many individuals to forgo future consumption to achieve more current consumption. In this case, the trade deficit should be viewed as immaterial since it merely reflects the free choices of the nation's people.

On the other hand, a trade surplus may correspond to prudent foreign saving and purchases of foreign productive assets, which may be used to support a growing retired population in the future. In this case, the trade surplus is a good thing for the nation. The trade surplus might also represent a period of repayment of past debt. This outcome may be acceptable if achieved together with growing living standards. However, if the surplus arises in a period of slow growth or falling GDP, then the surplus would correspond to painful reductions in living standards, which is clearly a bad outcome for the country. Finally, the trade surplus may occur as a result of the aggregated choices of many individuals who have acquired greater

past consumption by forgoing current consumption. In this case, the surplus should be viewed as immaterial to the nation as a whole.

KEY TAKEAWAYS

- Domestic spending measures the total value of purchases of goods and services in a country regardless of where the goods and services were produced. As such, it is a better way to measure the “consumption” in an economy affecting the nation’s standard of living as compared to “production” or GDP.
- When a country has a current account deficit, its national consumption exceeds its national production. When a country has a current account surplus, its national production exceeds its national consumption.
- Trade deficits become a problem over time if accumulated borrowings result in a substantial reduction in consumption and standard of living for its citizens during the repayment periods.
- The problems associated with a deficit occur not when the trade deficit is being run but in later periods when a trade surplus becomes necessary.
- Trade deficit problems are mitigated with GNP growth. The faster GNP grows, the lesser the decline in future consumption during the repayment period.

EXERCISES

1. Consider the Japanese economy over two periods of time: first period (today) and second period (the future). Suppose Japanese GDP today is \$2,000 billion (we'll use the U.S. dollar rather than the yen). Suppose Japan runs a current account surplus of 5 percent of GDP in the first period and lends money at the market interest rate of 5 percent.
 - a. What is the value of domestic spending on C , I , and G in the first period?
 - b. What would be the value of domestic spending in Japan in the second period if all the first period loans are repaid with interest and no economic growth occurs between periods?
2. Consider the following situations describing the actions of an individual household. Explain whether each situation is analogous to a country running a trade deficit, a trade surplus, or neither. Briefly explain why.
 - a. A student takes out a bank loan to finance a spring break vacation.
 - b. A family sells an antique watch to finance a purchase of 100 shares of a “hot” stock.
 - c. A retired couple cashes in a portion of their savings to finance their daily living expenses.
 - d. A carpenter builds a deck for a dentist in exchange for dental checkups for his kids.
 - e. A family pays off the last \$3,000 of its student loans.
3. Suppose that each situation listed is the dominant effect on a country's balance of payments. Indicate by filling in the blank spaces whether the current account and capital account will be in *surplus* or *deficit*.

	Current Account Balance	Financial Account Balance
a. A country is a net borrower from the rest of the world		

	Current Account Balance	Financial Account Balance
b. A country is repaying past debts		
c. A country exports more goods and services than it imports		
d. A country sells foreign assets and repatriates the proceeds		
e. A country is a net lender to the rest of the world		
f. A country earns more income on foreign assets than foreigners earn in its country		

14.4 Some Further Complications

LEARNING OBJECTIVE

1. Recognize how the long-term consequences of trade deficits change when they are financed by equity rather than debt.

The analysis of trade imbalances is further complicated by the fact that not all financial flows are debt obligations or IOUs (i.e., I owe you). In the previous stories, we assumed that all financial account transactions corresponded to international lending or borrowing. In actuality, many international asset transactions involve sales or purchases of productive assets. For example, if a foreigner purchases shares of Microsoft stock in the U.S. market, the transaction would be recorded as a credit entry on the financial account and would add to a financial account surplus.

However, in this case we could not claim that someone in the United States borrowed money from the rest of the world because there is no obligation to repay principal and interest in the future. Instead the foreign purchaser of the U.S. asset has purchased an ownership claim in a U.S. corporation that entitles him to the future stream of dividends plus capital gains if he sells the stock later at a higher price. If the company is profitable in the future, then the investors will earn a positive return. However, if the company suffers economic losses in the future, then the dividends may be discontinued and the stock's price may fall. Alternatively, the U.S. dollar could experience a significant depreciation. The end result could be losses for the foreign investor and a negative rate of return. In either case the foreign investor is not "entitled" to a return of his original investment or any additional return beyond. This same type of relationship arises for international real estate transactions and for foreign direct investment, which occurs when a foreign firm substantially owns and operates a company in another country.

To the extent that financial account flows correspond to asset purchases without repayment obligations, the stories above change somewhat. For example, suppose a country runs a trade deficit in period one and suppose further that the resulting financial account surplus corresponds to foreign purchases of U.S. real estate and businesses. In the first period, a country's standard of living could be raised above what is possible with balanced trade—not by borrowing money but by selling ownership claims on productive assets. In the second period, the country's standard of living need not be reduced since there is no repayment obligation.

This case is analogous to an individual who sells his watch at a pawnshop. In that period he is able to buy more than his income because he has divested some of his previously accumulated wealth. In the following period, he can once again make purchases equal to his income and thus need not suffer a reduction in his living standards.

The implication here is that nondebt asset flows may be less problematic than international loans because they do not require a reduction in living standards in the future. Of course, in this case, there is an additional concern that the country that sells off its assets may also be losing control of its productive assets and thus its citizens will not be the ones to earn positive returns on these domestic activities. This concern should be tempered for a few reasons. First, foreign-owned firms remain subject to the laws of the domestic country. Countries can prevent exorbitant profit taking by applying profit taxes. What's more, the foreign owners do not enjoy voting privileges and thus have less say over laws that might affect them. Second, foreign-owned firms generate employment opportunities for domestic citizens, and that serves to benefit the country. Finally, owners of firms, whether foreign or domestic, are generally motivated by similar desires—namely, to make the business successful—and successful businesses generally benefit the owners, the employees, and the consumers of the product.

As an example, consider the purchase in the 1980s of Rockefeller Center in New York City by a group of Japanese investors. Rockefeller Center is a centrally located building in New York City whose owners lease office space to businesses that wish to locate their offices there. Any owner of the building must compete with other businesses leasing office space throughout the city and thus must provide as high a quality and as low a price as possible. If the owners manage the property well and provide quality services, then they will have a lot of tenants and they will make a profit. If they provide poor services, then businesses will move out and the owners will lose money. Thus it really shouldn't matter to the tenants whether the owners are American or Japanese, only whether they are good managers of the office space. Similarly, the owners, regardless of nationality, will hire workers to maintain the facilities. These workers will benefit if the management is good and will suffer if it is poor, regardless of the owners' nationality. Finally, if the owners of the building are successful, then they deserve to earn a profit or return on their investment. If they provide poor services at high prices, then they will deserve to make a loss. Indeed, it shouldn't matter to anyone whether the owners are American or Japanese nationals.

KEY TAKEAWAY

- A trade deficit financed by sales of equity rather than debt does not require a repayment in the future or a subsequent decline in consumption. However, it does imply that a flow of the profits from domestic activities will accrue to foreigners rather than domestic residents.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The full form of the abbreviation “IOU.”
 - b. The terms representing the two broad types of assets; one is related to borrowing and lending and the other is related to ownership.
 - c. The type of asset represented by a bank certificate of deposit.
 - d. The type of asset represented by a common or preferred share of stock.
 - e. The type of asset represented by a checking account deposit.
 - f. The type of asset represented by the deed to a private golf course.
 - g. International purchases of this type of asset require repayment of principal and interest in the future.
 - h. International purchases of this type of asset do not require repayment of principal and interest.

14.5 How to Evaluate Trade Imbalances

LEARNING OBJECTIVE

1. Identify the conditions determining when a nation's trade imbalance is good, bad, or benign.

Review of Trade Imbalance Interpretations

A quick reading of business and financial newspapers and magazines often reveals a number of misunderstandings about economic relationships. One of the most notable is the widespread conviction that trade deficits are a troubling economic condition that indicates weakness in an economy, while trade surpluses are a sign of strength for an economy. Although these beliefs are well founded in some circumstances, they are not valid as a general principle. A careful look at the implications of trade imbalances reveals that trade deficits can, at times, be an indicator of rising economic stature, while trade surpluses can be associated with economic disaster. In many other cases, perhaps most, trade imbalances are simply benign—that is, they do not represent a serious threat or imply a notable benefit.

There are several reasons why misunderstandings about trade imbalances persist. The first problem relates to the terminology. A deficit, regardless of the context, sounds bad. To say that a business's books are in deficit, that a government's budget is in deficit, or that a country's trade balance is in deficit, simply sounds bad. A surplus, in contrast, sounds pretty good. For a business, clearly we'd prefer a surplus, to be in the black, to make a profit. Likewise, a budget surplus or a trade surplus must be good as well. Lastly, *balance* seems either neutral or possibly the ideal condition worth striving for. From an accountant's perspective, balance is often the goal. Debits must equal credits, and the books must balance. Surely, this terminology must contribute to the confusion, at least in a small way, but it is not accurate in describing trade imbalances in general.

A second reason for misunderstandings, especially with regard to deficits, may be a sense of injustice or inequity because foreigners are unwilling to buy as many of our goods as we buy of theirs. Fairness would seem to require reciprocity in international exchanges and therefore balanced trade. This misunderstanding could be easily corrected if only observers were aware that a country's balance of payments, which includes trade in goods, services, and assets, is always in balance. There are no unequal exchanges even when a country runs a trade deficit.

A third reason for the misunderstanding is that trade deficits are indeed bad for some countries in some situations while surpluses are sometimes associated with good economic outcomes. One needs only to note the many international debt crises experienced by countries after they had run persistent and very large trade deficits. One could also look at the very high growth rates of Japan in the 1980s and China in the last few decades for examples of countries with large trade surpluses that have seemingly fared very well.

However, despite these examples, one should not conclude that any country that has a trade deficit or whose trade deficit is rising is necessarily in a potentially dangerous situation; nor should we think that just because a country has a trade surplus that it is necessarily economically healthy. To see why, we must recognize that trade imbalances represent more than just an imbalance in goods and services trade.

Any imbalance in goods and services trade implies an equal and opposite imbalance in asset trade. When a country runs a trade deficit (more exhaustively labeled a *current account* deficit), it is also running a financial account surplus; similarly, a trade surplus corresponds to a financial account deficit. Imbalances on the financial account mean that a country is a net seller of international assets (if a financial account surplus) or a net buyer of international assets (if a financial account deficit).

One way to distinguish among good, bad, or benign trade imbalances is to recognize the circumstances in which it is good, bad, or benign to be a net international borrower or lender, a net purchaser, or seller of ownership shares in businesses and properties.

The International Investment Position

An evaluation of a country's trade imbalance should begin by identifying the country's net international asset or investment position. The investment position is like a balance sheet in that it shows the total holdings of foreign assets by domestic residents and the total holdings of domestic assets by foreign residents at a point in time. In the International Monetary Fund's (IMF) financial statistics, these are listed as domestic assets (foreign assets held by domestic residents) and domestic liabilities (domestic assets owned by foreign residents). In contrast, the financial account balance is more like an income statement that shows the changes in asset holdings during the past year. In other words, the international asset position of a country consists of stock variables while the financial account balance consists of flow variables.

A country's net international investment balance may either be in a debtor position, a creditor position, or in balance. If in a creditor position, then the value of foreign assets (debt and equity) held by domestic residents exceeds the value of domestic assets held by foreigners. Alternatively, we could say that domestic assets exceed domestic liabilities. If the reverse is true, so that domestic liabilities to foreigners exceed domestic assets, then the country would be called a debtor nation.

Asset holdings may consist of either debt obligations or equity claims. Debt consists of IOUs in which two parties sign a contract agreeing to an initial transfer of money from the lender to the borrower followed by a repayment according to an agreed schedule. The debt contract establishes an obligation for the borrower to repay principal and interest in the future. Equity claims represent ownership shares in potentially productive assets. Equity holdings do not establish obligations between parties, at least not in the form of guaranteed repayments. Once ownership in an asset is transferred from seller to buyer, all advantages and disadvantages of the asset are transferred as well.

Debt and equity obligations always pose several risks. The first risk with debt obligations is the risk of possible default (either total or partial). To the lender, default risk means that the IOU will not be repaid at all, that it will be repaid only in part, or that it is repaid over a much longer period than originally contracted. To the borrower, the risk of default is that future borrowing will likely become unavailable. In contrast, the advantage of default to the borrower is that not all the borrowed money is repaid.

The second risk posed by debt is that the real value of the repayments may be different than expected. This can arise because of unexpected inflation or unexpected currency value changes. Consider inflation first. If inflation is higher than expected, then the real value of debt repayment (if the nominal interest rate is fixed) will be lower than originally expected. This will be an advantage to the borrower (debtor), who repays less in real terms, and a disadvantage to the lender (creditor), who receives less in real terms. If inflation turns out to be less than expected, then the advantages are reversed.

Next, consider currency fluctuations. Suppose a domestic resident, who receives income in the domestic currency, borrows foreign currency in the international market. If the domestic currency depreciates, then the value of the repayments in domestic currency terms will rise even though the foreign currency repayment value remains the same. Thus currency depreciations can be harmful to borrowers of foreign currency. A similar problem can arise for a lender. Suppose a domestic resident purchases foreign currency and then lends it to a foreign resident (note in

this case the domestic resident is saving money abroad). Afterward, if the domestic currency appreciates, then foreign savings, once cashed in, will purchase fewer domestic goods and the lender will lose.

Similarly, various risks arise with equity purchases internationally because the asset's rate of return may turn out to be less than expected. This can happen for a number of different reasons. First, if the equity purchases are direct investment in a business, then the return on that investment will depend on how well the business performs. If the market is vibrant and management is good, then the investment will be profitable. Otherwise, the rate of return on the investment could be negative; the foreign investor could lose money. In this case, all the risk is borne by the investor, however. The same holds for stock purchases. Returns on stocks may be positive or negative, but it is the purchaser who bears full responsibility for the return on the investment. As with debt, equity purchases can suffer from exchange rate risk as well. When foreign equities are purchased, their rate of return in terms of domestic currency will depend on the currency value. If the foreign currency in which assets are denominated falls substantially in value, then the value of those assets falls along with it.

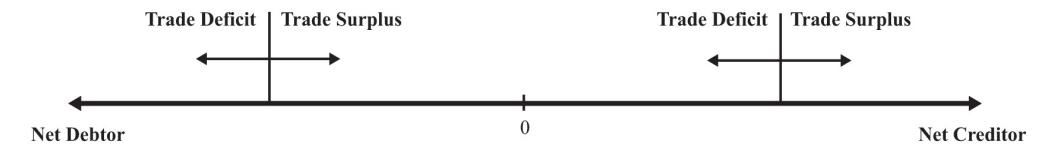
Four Trade Imbalance Scenarios

There are four possible situations that a country might face. It may be

1. a debtor nation with a trade deficit,
2. a debtor nation with a trade surplus,
3. a creditor nation with a trade deficit,
4. a creditor nation with a trade surplus.

Figure 14.6 "International Asset Positions" depicts a range of possible international investment positions. On the far left of the image, a country would be a net debtor nation, while on the far right, it would be a net creditor nation. A trade deficit or surplus run in a particular year will cause a change in the nation's asset position assuming there are no capital gains or losses on net foreign investments. A trade deficit would generally cause a leftward movement in the nation's investment position implying either a reduction in its net creditor position or an increase in its net debtor position. A trade surplus would cause a rightward shift in a country's investment position implying either an increase in its net creditor position or a decrease in its net debtor position.

Figure 14.6 International Asset Positions



An exception to this rule occurs whenever there are changes in the market value of foreign assets and when the investment position is calculated using current market values rather than original cost. For example, suppose a country has balanced trade in a particular year and is a net creditor nation. If the investment position is evaluated using original cost, then since the current account is balanced, there would be no change in the investment position. However, if the investment position is evaluated at current market values, then the position can change even with balanced trade. In this case, changes in the investment position arise due to capital gains or losses. Real estate or property valuations may change, portfolio investments in stock markets may rise or fall, and currency value changes may also affect the values of national assets and liabilities.

The pros and cons of a national trade imbalance will depend on which of the four situations describes the current condition of the country. We'll consider each case in turn next.

Case 1: Net Debtor Nation Running a Current Account Deficit

This is perhaps the most common situation in the world, or at least this type of case gets the most attention. The main reason is that large trade deficits run persistently by countries, which are also large debtor nations, can eventually be unsustainable. Examples of international debt crises are widespread. They include the third world debt crisis of the early 1980s, the Mexican crisis in 1994, and the Asian crisis in 1997.

However, not all trade deficits nor all debtor countries face eventual default or severe economic adjustment. Indeed, for some countries, a net debtor position with current account deficits may be an ideal economic situation. To distinguish the good cases from the bad requires us to think about situations in which debt is good or bad.

As mentioned earlier, a current account deficit means that a country is able to spend more on goods and services than it produces during the year. The additional spending can result in increases in consumption, investment, and/or government spending. The country accomplishes this as a net debtor country by borrowing from

the rest of the world (incurring debt), or by selling some of its productive assets (equities).

Let's consider a few scenarios.

First, suppose the current account deficit is financed by borrowing money from the rest of the world (i.e., incurring debt). Suppose the additional spending over income is on consumption and government goods and services. In this case, the advantage of the deficit is that the country is able to consume more private and public goods while it is running the deficit. This would enhance the nation's average standard of living during the period the deficit is being run. The disadvantage is that the loans that finance the increase in the standard of living must be repaid in the future. During the repayment period, the country would run a current account surplus, resulting in national spending below national income. This might require a reduction in the country's average standard of living in the future.

This scenario is less worrisome if the choices are being made by private citizens. In this case, individuals are freely choosing to trade off future consumption for current consumption. However, if the additional spending is primarily on government goods and services, then it will be the nation's taxpayers who will be forced to repay government debt in the future by reducing their average living standards. In other words, the future taxpayers' well-being will be reduced to pay for the extra benefits accruing to today's taxpayers.

Possible reductions in future living standards can be mitigated or eliminated if the economy grows sufficiently fast. If national income is high enough in the future, then average living standards could still rise even after subtracting repayment of principal and interest. Thus trade deficits are less worrisome when both current and future economic growth are more rapid.

One way to stimulate economic growth is by increasing spending on domestic investment. If the borrowed funds that result when a country runs a current account deficit are used for investment rather than consumption or if the government spending is on infrastructure, education, or other types of human and physical capital, then the prospects for economic growth are enhanced.

Indeed, for many less-developed countries and countries in transition from a socialist to capitalist market, current account deficits represent potential salvation rather than a curse. Most poor countries suffer from low national savings rates (due to low income) and inadequate tax collection systems. One obvious way to finance investment in these countries is by borrowing from developed countries that have

much higher national savings rates. As long as the investments prove to be effective, much more rapid economic growth may be possible.

Thus trade deficits for transitional and less-developed economies are not necessarily worrisome and may even be a sign of strength if they are accompanied by rising domestic investment and/or rising government expenditures on infrastructure.

The main problem with trade deficits arises when they result in a very large international debt position. (Arguably, one could claim that international debt greater than 50 percent of GDP is very large.) In this circumstance, it can lead to a crisis in the form of a default on international obligations. However, the international debt position figures include both debt and equities, and only the debt can be defaulted on. Equities, or ownership shares, may yield positive or negative returns but do not represent the same type of contractual obligations. A country would never be forced to repay foreign security holders for its losses simply because its value on the market dropped. Thus a proper evaluation of the potential for default should only look at the net international “debt” position after excluding the net position on equities.

Default becomes more likely the larger the external debt relative to the countries’ ability to repay. Ability to repay can be measured in several ways. First, one can look at net debt relative to GDP. Since it measures annual national income, GDP represents the size of the pool from which repayment of principal and interest is drawn—the larger the pool, the greater the ability of the country to repay. Alternatively, the lower the country’s net debt to GDP ratio, the greater the country’s ability to repay.

A second method to evaluate ability to repay is to consider net debt as a percentage of exports of goods and services. This is especially relevant when international debt is denominated in foreign currencies. In this case, the primary method to acquire foreign currencies to make repayment of debt is through the export of goods and services. (The alternative method is to sell domestic assets.) Thus the potential for default may rise if the country’s ratio of net external debt to exports is larger.

Notice, though, that the variable to look at to evaluate the risk of default is the net debt position, *not* the trade deficit. The trade deficit merely reveals the change in the net debt position during the past year and does not record total outstanding obligations. In addition, a trade deficit can be run even while the net “debt” position falls. This could occur if the trade deficit is financed primarily with net equity sales rather than net debt obligations. Thus the trade deficit, by itself, does not reveal a complete picture regarding the potential for default.

Next, we should consider what problems are associated with default. Interestingly, it is not really default itself that is immediately problematic but the actions taken to avoid default. If default on international debt does occur, international relationships with creditor countries would generally suffer. Foreign banks that are not repaid on past loans will be reluctant to provide loans in the future. For a less-developed country that needs foreign loans to finance productive investment, these funds may be cut off for a long period and thus negatively affect the country's prospects for economic growth. On the positive side, default is a benefit for the defaulting country in the short-run since it means that borrowed funds are not repaid. Thus the country enjoys the benefits of greater spending during the previous periods when trade deficits are run but does not have to suffer the consequences of debt repayment. With regard to the country's international debt position, default would cause an immediate discrete reduction in the country's debt position.

The real problem arises when economic shocks suddenly raise external obligations on principal and interest, making a debt that was once sustainable suddenly unsustainable. In these cases, it is the effort made to avoid default that is the true source of the problem.

Inability to repay foreign debt arises either if the value of payments suddenly increases or if the income used to finance those payments suddenly falls. Currency depreciations are a common way in which the value of repayments can suddenly rise. If foreign debt is denominated in foreign currency, then domestic currency depreciation implies an appreciation in the value of external debt. If the currency depreciation is large enough, a country may become suddenly unable to make interest and principal repayments. Note, however, that if external debt were denominated in domestic currency, then the depreciation would have no effect on the value of interest and principal repayments. This implies that countries with large external debts are in greater danger of default if (1) their currency value is highly volatile and (2) the external debt is largely denominated in foreign currency.

A second way in which foreign interest obligations can suddenly rise is if the obligations have variable interest rates and if the interest rates suddenly rise. This was one of the problems faced by third world countries during the debt crisis in the early 1980s. Loans received from the U.S. and European banks carried variable interest rates to reduce the risk to the banks from unexpected inflation. When restrictive monetary policy in the United States pushed up U.S. interest rates, interest obligations by foreign countries also suddenly rose. Thus international debt with variable interest rates potentially raises the likelihood of default.

Default can also occur if a country's ability to repay suddenly falls. This can occur if the country enters into a recession. Recessions imply falling GDP, which reduces the pool of funds available for repayment. If the recession is induced by a reduction in exports, perhaps because of recessions in major trading partner countries, then the ability to finance foreign interest and principal repayments is reduced. Thus a recession in the midst of a large international debt position can risk potential default on international obligations.

But what are the problems associated with a sudden increase in debt repayment if default on the debt does not occur? The problem, really, is that the country might suddenly have to begin running current account surpluses to maintain repayments of its international obligations. Remember that trade deficits mean that the country can spend more than its income. By itself, that's a good thing. Current account surpluses, though, mean that the country must spend less than its income. That's the bad thing, especially if it occurs in the face of an economic recession.

Indeed, this is one of the problems the U.S. economy is facing in the midst of the current recession. As the U.S. GDP began to fall in the fall of 2008, the U.S. trade deficit also fell. For the "trade deficits are bad" folks, this would seem to be a good thing. However, it really indicated that not only was U.S. production falling but, because its trade deficit was also falling, its consumption was falling even faster. In terms of standard of living, the drop in the U.S. trade deficit implied a worsening of the economic conditions of its citizens.

However, since this problem arises only when a net debtor country runs a current account surplus, we'll take up this case in the next section. Note well though that the problems associated with a trade deficit run by a net debtor country are generally not visible during the period in which the trade deficit is run. It is more likely that a large international debt will pose problems in the future if or when substantial repayment begins.

In summary, the problem of trade deficits run by a net debtor country is more worrisome

1. the larger the net debtor position,
2. the larger the net debt (rather than equity) position,
3. the larger the CA deficit (greater than 5 percent of GDP is large according to some, although large deficit with small net debtor position is less worrisome),
4. the more net debt is government obligations or government backed,
5. the larger the government deficit,

6. if a high percentage of debt is denominated in foreign currency and if the exchange rate has or will depreciate substantially,
7. if rising net debt precedes slower GDP growth,
8. if rising net debt correlates with falling investment,
9. if deficits correspond to “excessive” increase in $(C + G)$ per capita (especially if G is not capital investment),
10. if interest rate on external debt is variable,
11. if a large recession is imminent.

The situation is benign or beneficial if the reverse occurs.

Case 2: Net Debtor Nation Running a Current Account Surplus

This case generally corresponds to a country in the process of repaying past debt. Alternatively, foreigners may be divesting themselves of domestic equity assets (i.e., selling previously purchased equities, like stocks and real estate, back to domestic residents). In either case, the trade surplus will reduce the country’s net debtor position and will require that domestic spending is less than national income. This case is especially problematic if it arises because currency depreciation has forced a sudden change in the country’s required repayments on international debt. This is the outcome when a series of trade deficits proves to be unsustainable. What unsustainability means is that the deficits can no longer be continued. Once external financing is no longer available, the country would not have the option to roll over past obligations. In this case, in the absence of default, the country’s net repayment on current debt would rise and push the financial account into deficit and hence the trade account into surplus.

When this turnaround occurs rapidly, the country suddenly changes from a state in which it spends more on consumption, investment, and government than its income to a state in which it spends less on these items than its income. Even if GDP stayed the same, the country would suffer severe reductions in its standard of living and reductions in its investment spending. The rapid reduction in domestic demands is generally sufficient to plunge the economy into a recession as well. This reduction in GDP further exacerbates the problem.

This problematic outcome is made worse nationally when most of the debt repayment obligations are by the domestic government or if the external obligations are government-backed. A government that must suddenly make larger than expected repayments of debt must finance it either by raising taxes or by reducing government benefits. The burden of the repayment is then borne by the general population because it must all come from taxpayers. Exactly who suffers more or less will depend on the nature of the budget adjustments, although it often

seems that poorer segments of the population bear the brunt of the adjustment costs.

If the sudden increase in debt repayment were primarily by private firms, then the burdens would fall on the associates of those firms rather than the general population. If this occurs on a small scale, we can view this as normal adjustments in a free market system: some firms always go bust, forcing dislocations of labor and capital. The general population in this case would not bear the burden of adjustment unless they are affiliated with the affected firms.

However, even if the debt repayment burden is private and even if the government had not previously guaranteed that debt, the government may feel compelled to intervene with assistance if many private firms are negatively affected. This will perhaps be even more likely if the affected private debt is held by major national banks. Default by enough banks can threaten the integrity of the banking system. Government intervention to save the banks would mean that the general population would essentially bear the burdens of private mistakes.

This kind of rapid reversal is precisely what happened to Indonesia, Thailand, Malaysia, and South Korea in the aftermath of the Asian currency crisis in 1997. Afterward, these countries recorded substantial current account surpluses. These surpluses should not be viewed as a sign of strong vibrant economies; rather, they reflect countries that are in the midst of recessions, struggling to repay their past obligations, and that are now suffering a reduction in average living standards as a consequence.

The most severe consequences of a current account surplus as described above arise when the change from trade deficit to surplus is abrupt. If, on the other hand, the transition is smooth and gradual, then the economy may not suffer noticeably at all. For example, consider a country that has financed a period of extra spending on infrastructure and private investment by running trade deficits and has become a net international debtor nation. However, once the investments begin to take off, fueling rapid economic growth, the country begins to repay more past debt than the new debt that it incurs each period. In this case, the country could make a smooth transition from a trade deficit to a trade surplus. As long as GDP growth continued sufficiently fast, the nation might not even need to suffer reductions in its average living standards even though it is spending less than its income during the repayment period.

In summary, the situation of a net debtor nation running current account surpluses is more worrisome if

1. surpluses follow default,
2. GDP growth rate is low or negative,
3. the investment rate is low or falling,
4. real $C + G$ per capita is falling,
5. surplus corresponds to rising net debt and larger equity sales.

The situation is benign or beneficial if the reverse occurs.

Case 3: Net Creditor Nation Running a Current Account Surplus

A net creditor country with trade surpluses is channeling savings to the rest of the world either through lending or through the purchase of foreign productive assets. The situation is generally viewed as prudent but may have some unpleasant consequences. Recall that a country with a trade surplus is spending less on consumption, investment, and government combined than its national income. The excess is being saved abroad. Net creditor status means that the country has more total savings abroad than foreigners have in their country.

The first problem may arise if the surplus corresponds to the substitution of foreign investment for domestic investment. In an era of relatively free capital mobility, countries may decide that the rate of return is higher and the risk is lower on foreign investments compared to domestic investments. If domestic investment falls as a result, future growth prospects for the country are reduced as well. This situation has been a problem in Russia and other transition economies. As these economies increased their private ownership of assets, a small number of people became extremely wealthy. In a well-functioning economy with good future business prospects, wealth is often invested internally helping to fuel domestic growth. However, in many transition economies, wealth holders decided that it was too risky to invest domestically because uncertainty about future growth potential was very low. So instead, they saved their money abroad, essentially financing investment in much healthier and less risky economies.

China is another creditor country running a trade surplus today. It is, however, in a different situation than Russia or the transition economies in the 1990s. China's internal investment rate is very high and its growth rate has been phenomenal over the past twenty years or more. The fact that it has a trade surplus means that as a nation it is saving even more than necessary to finance its already high investment levels. The excess it is lending abroad, thereby raising its international creditor position. (See [Chapter 12 "Introductory Finance Issues: Current Patterns, Past History, and International Institutions"](#) for more details.) If it was to redirect that saving domestically, it may not be able to fuel additional growth since their investment spending is already so high. Their trade surplus also means that its

average standard of living is well below what is possible because it is saving the surplus abroad rather than spending it on consumption or government goods at home.

A second problem arises even if domestic investment remains high. With domestic investment kept high, the cost of the large surpluses must be felt as a reduction in consumption and government spending. In this case, a large trade surplus leads to a reduction in average living standards for the country. This is a point worth emphasizing. Countries that run trade surpluses suffer a reduction in living standards, not an increase, relative to the case of balanced trade.

Another potential problem with being a net creditor country is the risk associated with international lending and asset purchases. First of all, foreign direct investments may not pay off as hoped or expected. Portfolio investments in foreign stock markets can suddenly be reduced in value if the foreign stock market crashes. On international loans, foreign nations may default on all or part of the outstanding loans, may defer payments, or may be forced to reschedule payments. This is a more likely event if the outstanding loans are to foreign countries with national external debts that may prove unsustainable. If the foreign country suffers rapid currency depreciation and if the foreign loans are denominated in domestic currency, then the foreign country may be forced to default. Defaults may also occur if the foreign debtor countries suffer severe recessions. The creditor nation in these cases is the one that must suffer the losses.

It was this situation that was especially serious for the United States at the onset of the third world debt crisis in the early 1980s. At that time, a number of large U.S. banks had a considerable proportion of their asset portfolios as loans to third world countries. Had these countries defaulted en masse, it would have threatened the solvency of these banks and could have led to a serious banking crisis in the U.S. economy.

Alternatively, suppose the surplus country has made external loans in the foreign countries' currency. If the foreign currency depreciates, even if only gradually, then the value of the foreign assets falls in terms of the domestic currency. The realized rates of return on these assets could then become negative, falling far short of returns on comparable domestic assets.

This is the dilemma that China faces today. The Chinese government has accumulated almost \$1 trillion of U.S. Treasury bonds as a result of its persistent current account surpluses over the past decade. All of this debt is denominated in U.S. dollars, making it subject to exchange rate risk. If the Chinese relent to U.S. pressure to allow their fixed currency value appreciate to the U.S. dollar, then the

value of these U.S. assets falls in value and reduces their future returns. The Chinese are also worried about the potential for future U.S. inflation due to the expansionary monetary policy used during the current economic crisis. If inflation does arise in the future, the value of the trillion dollars of foreign debt would also be reduced. This situation is epitomized with a popular parable that says, “If you owe me a thousand dollars, then *you* have a problem, but if you owe me a million dollars, then *I* have a problem.” Even though the United States is the debtor and the Chinese the creditor, the Chinese now have a problem because they may have lent too much to the United States.

In summary, the situation of a net creditor nation running current account surpluses is worrisome if the

1. net credit position is very large,
2. current account surplus is very large,
3. GDP growth rate is low,
4. investment rate is low or falling,
5. $C + G$ per capita is low or falling,
6. surplus involves lending denominated in a foreign currency that may afterward depreciate,
7. domestic currency has appreciated substantially,
8. foreign asset values have fallen substantially.

The situation is benign or beneficial if the reverse occurs.

Case 4: Net Creditor Nation Running a Current Account Deficit

In general, a deficit run by a country that is a net creditor is least likely to be problematic. Essentially, this describes a country that is drawing down previously accumulated savings. The deficit also implies that the country is spending more than its income. This situation is especially good if it allows the country to maintain living standards during a recession. This case would also be good if a country with a rapidly aging population is drawing down previous savings to maintain average living standards.

The current account deficit can cause problems if as in case one, the deficit corresponds to falling investment and increases in consumption and government expenditures. If these changes occur while the economy continues to grow, then it may indicate potential problems for future economic growth. However, if the same changes occur while the economy is in a recession, then the effect would be to maintain average living standards by drawing down external savings. If this occurs only during the recession, then the long-term effect on growth would be mitigated.

This case can be a problem if the net creditor position is extremely large. A large amount of foreign savings can always potentially drop in value given currency fluctuations as described above in case three. However, the current account deficit only serves to reduce this potential problem since it reduces the country's net creditor position.

In summary, the situation of a net creditor nation running current account deficits is worrisome if

1. the net creditor status is smaller and the deficit is larger (although this is generally less worrisome than if the country were a net debtor),
2. investment is falling (although a temporary drop in investment is likely in a recession),
3. $C + G$ per capita is rising rapidly.

The situation is benign or beneficial if the reverse occurs.

KEY TAKEAWAYS

- Since trade deficits are not always bad and trade surpluses not always good, it is important to know how to judge a country's trade imbalance.
- Trade deficits are more worrisome when a country is a large international debtor and when growth or prospective growth is low.
- Trade deficits are less worrisome if international debt is low or if the country is a creditor nation.
- Trade deficits are less worrisome if they accompany increased investment and other stimuli to economic growth.
- Trade surpluses are more worrisome when the foreign credits reduce domestic investment sufficiently to reduce growth.
- Trade surpluses are more worrisome when future repayments will likely be lower than anticipated. This can occur if the credits are exceedingly large or denominated in foreign currency.
- Trade surpluses are more worrisome when they arise suddenly in association with a large international debtor position.

EXERCISES

1. Suppose the hypothetical country of Avalon has a current account deficit of \$20 billion this year. From the two scenarios listed in each part below, identify which scenario would make this deficit more worrisome to an economic analyst and which scenario would be less worrisome. Briefly explain why.

- a. Scenario 1: Avalon's GDP is \$80 billion dollars per year.

Scenario 2: Avalon's GDP is \$800 billion per year.

- b. Scenario 1: Avalon is a net debtor country.

Scenario 2: Avalon is a net creditor country.

- c. Scenario 1: Avalon's annual consumption spending is 50 percent of GDP.

Scenario 2: Avalon's annual consumption spending is 90 percent of GDP.

- d. Scenario 1: Avalon's GDP grew 1 percent last year.

Scenario 2: Avalon's GDP grew 10 percent last year.

2. Below are the economic data for five fictitious countries running trade deficits. Dollar amounts are in billions, and percentages are relative to GDP.

	Alpha (%)	Beta (%)	Gamma (%)	Delta (%)	Epsilon (%)
GDP	\$260	\$340	\$135	\$400	\$840
Trade Deficit (TD)	9.1	9.7	2.5	5.7	6.0

	Alpha (%)	Beta (%)	Gamma (%)	Delta (%)	Epsilon (%)
Projected GDP Growth	+2.0	+10.2	+3.0	+1.0	+5.5
Net International Investment Position (IIP)	75 debtor	30 creditor	20 debtor	60 debtor	5 debtor
Domestic Investment (I)	18	35	16	13	27

Suppose you work for the International Monetary Fund, and it has asked you to assess which two of these five countries' trade deficits are most likely to pose future repayment problems. Provide a brief explanation justifying your assessment.

3. Consider the fictitious country of Malamar. Economic data for Malamar are presented in the table below. Note that Malamar is currently running a trade deficit of \$60 billion.

Trade Deficit (TD)	\$60 billion
GDP	\$1,000 billion
GDP Growth—Past 3 Years (Growth -)	-1.2%
Projected GDP Growth—Next 3 Years (Growth +)	8.5%
Net International Investment Position (IIP)	-\$800 billion (debtor)
Domestic Investment (I)	\$350 billion

In the table below, reference the above data (either directly or in combination) in the first column and indicate in the second column whether this information tends to make Malamar's deficit *more* worrisome or *less* worrisome. One example is provided to illustrate.

Data	More or Less Worrisome
$TD/GDP = 6\% \text{ percent}$	More

Data	More or Less Worrisome

4. Consider the following statements concerning current account balances. Explain in what sense, if any, the statements are valid. In what sense, if any, are the statements misguided?
- A current account deficit implies that our nation is giving away money to the rest of the world.
 - A current account deficit indicates that a country has exported jobs to the rest of the world.
 - A current account deficit implies that the nation must have a reduced standard of living in the future.

Chapter 15

Foreign Exchange Markets and Rates of Return

People trade one national currency for another for one reason: they want to do something with the other currency. What they might do consists of one of two things: either they wish to spend the money, acquiring goods and services, or they wish to invest the money.

This chapter introduces the foreign exchange market for currency trades. It highlights some of the more obvious, although sometimes confusing, features and then turns attention to the motivations of foreign investors. One of the prime motivations for investing in another country is because one hopes to make more money on an investment abroad. How an investor calculates and compares those rates of returns are explored in this chapter.

15.1 The Forex: Participants and Objectives

LEARNING OBJECTIVE

1. Learn who participates in foreign exchange markets and why.

The foreign exchange market (Forex) is not a market like the New York Stock Exchange, where daily trades of stock are conducted in a central location. Instead, the Forex refers to the activities of major international banks that engage in currency trading. These banks act as intermediaries between the true buyers and sellers of currencies (i.e., governments, businesses, and individuals). These banks will hold foreign currency deposits and stand ready to exchange these for domestic currency upon demand. The exchange rate (ER) will be determined independently by each bank but will essentially be determined by supply and demand in the market. In other words, the bank sets the exchange rate at each moment to equalize its supply of foreign currency with the market demand. Each bank makes money by collecting a transactions fee for its “exchange services.”

It is useful to categorize two distinct groups of participants in the Forex, those whose transactions are recorded on the current account (importers and exporters) and those whose transactions are recorded on the financial account (investors).

Importers and Exporters

Anyone who imports or exports goods and services will need to exchange currencies to make the transactions. This includes tourists who travel abroad; their transactions would appear as services in the current account. These businesses and individuals will engage in currency trades daily; however, these transactions are small in comparison to those made by investors.

International Investors, Banks, Arbitrageurs, and Others

Most of the daily currencies transactions are made by investors. These investors, be they investment companies, insurance companies, banks, or others, are making currency transactions to realize a greater return on their investments or holdings. Many of these companies are responsible for managing the savings of others. Pension plans and mutual funds buy and sell billions of dollars worth of assets daily. Banks, in the temporary possession of the deposits of others, do the same. Insurance companies manage large portfolios that act as their capital to be used to

pay off claims on accidents, casualties, and deaths. More and more of these companies look internationally to make the most of their investments.

It is estimated by the Bank of International Settlements that over \$3 trillion (or \$3,000 billion) worth of currency is traded every day. Only about \$60 to \$100 billion of trade in goods and services takes place daily worldwide. This suggests that many of the currency exchanges are done by international investors rather than importers and exporters.

Investment Objectives

Investors generally have three broad concerns when an investment is made. They care about how much money the investment will earn over time, they care about how risky the investment is, and they care about how liquid, or convertible, the asset is.

1. **Rate of return (RoR)**¹. The percentage change in the value of an asset over some period.

Investors purchase assets as a way of saving for the future. Anytime an asset is purchased, the purchaser is forgoing current consumption for future consumption. To make such a transaction worthwhile the investors hope (sometimes expect) to have more money for future consumption than the amount they give up in the present. Thus investors would like to have as high a rate of return on their investments as possible.

Example 1: Suppose a Picasso painting is purchased in 1996 for \$500,000. One year later, the painting is resold for \$600,000. The rate of return is calculated as

$$\frac{(600,000 - 500,000)}{500,000} \times 100 = \frac{100,000}{500,000} \times 100 = 0.20 \times 100 = 20\%.$$

Example 2: \$1,000 is placed in a savings account for one year at an annual interest rate of 10 percent. The interest earned after one year is $\$1,000 \times 0.10 = \100 . Thus the value of the account after one year is \$1,100. The rate of return is

$$\frac{1100 - 1000}{1000} \times 100 = \frac{100}{1000} \times 100 = 0.10 \times 100 = 10\%.$$

1. The percentage change in the value of an asset over some period.

This means that the rate of return on a domestic interest-bearing account is merely the interest rate.

2. **Risk.** The second primary concern of investors is the riskiness of the assets. Generally, the greater the expected rate of return, the greater the risk. Invest in an oil wildcat endeavor and you might get a 1,000 percent return on your investment—that is, if you strike oil. The chances of doing so are likely to be very low, however. Thus a key concern of investors is how to manage the trade-off between risk and return.
3. **Liquidity.** Liquidity essentially means the speed with which assets can be converted to cash. Insurance companies need to have assets that are fairly liquid in the event that they need to pay out a large number of claims. Banks also need to be able to make payouts to their depositors, who may request their money back at any time.

KEY TAKEAWAYS

- Participants in the foreign exchange markets can be classified into traders and investors.
- Traders export or import goods and services whose transactions appear on the current account of the balance of payments.
- Investors purchase or sell assets whose transactions appear on the financial account of the balance of payments.
- The three main concerns for any investor are first to obtain a high rate of return, second to minimize the risk of default, and third to maintain an acceptable degree of liquidity.
- The rate of return on an asset is the percentage change in its value over a period.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This group enters the foreign exchange market to make transactions that will be recorded on the current account.
 - b. This group enters the foreign exchange market to make transactions that will be recorded on the financial account.
 - c. The percentage change in the value of an asset over some period.
 - d. The term used to describe the ease with which an asset can be converted to cash.
 - e. The term used to describe the possibility that an asset will not return what is originally expected.
 - f. A list of three main objectives for international investors.
 - g. The rate of return on a share of stock whose value rises during the year from \$5.50 per share to \$6.50 per share.
 - h. The rate of return on a commercial office building that was purchased one year ago for \$650,000 and sold today for \$600,000.

15.2 Exchange Rate: Definitions

LEARNING OBJECTIVE

1. Learn some of the basic definitions regarding currency markets and exchange rates.

Anyone who has ever traveled to another country has probably had to deal with an exchange rate between two currencies. (I say “probably” because a person who travels from, say, Italy to Spain continues to use euros.) In a sense, exchange rates are very simple. However, despite their simplicity they never fail to generate confusion. To overcome that confusion this chapter begins by offering straightforward definitions and several rules of thumb that can help with these problems.

The **exchange rate (ER)**² represents the number of units of one currency that exchanges for a unit of another. There are two ways to express an exchange rate between two currencies (e.g., between the U.S. dollar [\$] and the British pound [£]). One can either write $\$/\text{£}$ or $\text{£}/\$$. These are reciprocals of each other. Thus if E is the $\$/\text{£}$ exchange rate and V is the $\text{£}/\$$ exchange rate, then $E = 1/V$.

For example, on January 6, 2010, the following exchange rates prevailed:

$$E\$/\text{£} = 1.59, \text{ which implies } V\text{£}/\$ = 0.63,$$

and

$$V\text{¥}/\$ = 92.7, \text{ which implies } E\$/\text{¥} = 0.0108.$$

Currency Value

It is important to note that the value of a currency is always given in terms of another currency. Thus the value of a U.S. dollar in terms of British pounds is the $\text{£}/\$$ exchange rate. The value of the Japanese yen in terms of dollar is the $\$/\text{¥}$ exchange rate.

2. Represents the number of units of one currency that exchanges for a unit of another.

Note that we always express the value of all items in terms of something else. Thus the value of a quart of milk is given in dollars, not in quarts of milk. The value of car

is also given in dollar terms, not in terms of cars. Similarly, the value of a dollar is given in terms of something else, usually another currency. Hence, the rupee/dollar exchange rate gives us the value of the dollar in terms of rupees.

This definition is especially useful to remember when one is dealing with unfamiliar currencies. Thus the value of the euro (€) in terms of British pounds is given as the $\text{£}/\text{€}$ exchange rate.

Similarly, the peso/euro exchange rate refers to the value of the euro in terms of pesos.

Currency appreciation³ means that a currency *appreciates* with respect to another when *its value rises* in terms of the other. The dollar appreciates with respect to the yen if the $\text{¥}/\text{\$}$ exchange rate rises.

Currency depreciation⁴, on the other hand, means that a currency *depreciates* with respect to another when *its value falls* in terms of the other. The dollar depreciates with respect to the yen if the $\text{¥}/\text{\$}$ exchange rate falls.

Note that if the $\text{¥}/\text{\$}$ rate rises, then its reciprocal, the $\text{\$}/\text{¥}$ rate, falls. Since the $\text{\$}/\text{¥}$ rate represents the value of the yen in terms of dollars, this means that when the dollar appreciates with respect to the yen, the yen must depreciate with respect to the dollar.

The rate of appreciation (or depreciation) is the percentage change in the value of a currency over some period.

Example 1: U.S. dollar (US\$) to the Canadian dollar (C\$)

On January 6, 2010, $E_{\text{C\$}}/\text{US\$} = 1.03$.

On January 6, 2009, $E_{\text{C\$}}/\text{US\$} = 1.19$.

Use the percentage change formula, (new value - old value)/old value:

- 3. A currency *appreciates* with respect to another when *its value rises* in terms of the other.
- 4. A currency *depreciates* with respect to another when *its value falls* in terms of the other.

$$\frac{(1.03 - 1.19)}{1.19} = \frac{-0.16}{1.19} = -0.134.$$

Multiply by 100 to write as a percentage to get

$$-0.134 \times 100 = -13.4\%.$$

Since we have calculated the change in the value of the U.S. dollar in terms of Canadian dollar, and since the percentage change is negative, this means that the dollar has depreciated by 13.4 percent with respect to the C\$ during the previous year.

Example 2: U.S. dollar (\$) to the Pakistani rupee (R)

On January 6, 2010, $E_{R/\$} = 84.7$.

On January 6, 2010, $E_{R/\$} = 79.1$.

Use the percentage change formula, (new value – old value)/old value:

$$\frac{(84.7 - 79.1)}{79.1} = + \frac{5.6}{79.1} = +0.071.$$

Multiply by 100 to write as a percentage to get

$$+0.071 \times 100 = +7.1\%.$$

Since we have calculated the change in the value of the U.S. dollar, in terms of rupees, and since the percentage change is positive, this means that the dollar has appreciated by 7.1 percent with respect to the Pakistani rupee during the past year.

Other Exchange Rate Terms

- 5. The process of buying a product when its price is low and then reselling it after its price rises to make a profit.
- 6. The exchange rate that prevails *on the spot*, that is, for trades to take place immediately.
- 7. The rate that appears on a contract to exchange currencies either 30, 60, 90, or 180 days in the future.

Arbitrage⁵ generally means buying a product when its price is low and then reselling it after its price rises in order to make a profit. Currency arbitrage means buying a currency in one market (e.g., New York) at a low price and reselling, moments later, in another market (e.g., London) at a higher price.

The **spot exchange rate**⁶ refers to the exchange rate that prevails *on the spot*, that is, for trades to take place immediately. (Technically, it is for trades that occur within two days.)

The **forward exchange rate**⁷ refers to the rate that appears on a contract to exchange currencies either 30, 60, 90, or 180 days in the future.

For example, a corporation might sign a contract with a bank to buy euros for U.S. dollars sixty days from now at a predetermined ER. The predetermined rate is called the sixty-day forward rate. Forward contracts can be used to reduce exchange rate risk.

For example, suppose an importer of BMWs is expecting a shipment in sixty days. Suppose that upon arrival the importer must pay €1,000,000 and the current spot ER is 1.20 \$/€.

Thus if the payment were made today it would cost \$1,200,000. Suppose further that the importer is fearful of a U.S. dollar depreciation. He doesn't currently have the \$1,200,000 but expects to earn more than enough in sales over the next two months. If the U.S. dollar falls in value to, say, 1.30 \$/€ within sixty days, how much would it cost the importer in dollars to purchase the BMW shipment?

The shipment would still cost €1,000,000. To find out how much this is in dollars, multiply €1,000,000 by 1.30 \$/€ to get \$1,300,000.

Note that this is \$100,000 more for the cars simply because the U.S. dollar value changed.

One way the importer could protect himself against this potential loss is to purchase a forward contract to buy euros for U.S. dollars in sixty days. The ER on the forward contract will likely be different from the current spot ER. In part, its value will reflect market expectations about the degree to which currency values will change in the next two months. Suppose the current sixty-day forward ER is 1.25 \$/€, reflecting the expectation that the U.S. dollar value will fall. If the importer purchases a sixty-day contract to buy €1,000,000, it will cost him \$1,250,000 (i.e., $\$1,000,000 \times 1.25 \text{ \$/\euro}$). Although this is higher than what it would cost if the exchange were made today, the importer does not have the cash available to make the trade today, and the forward contract would protect the importer from an even greater U.S. dollar depreciation.

8. When the forward exchange rate is such that a forward trade costs more (or buys less foreign currency) than a trade on the spot market today.

9. When the forward exchange rate is such that a forward trade costs less (or buys more foreign currency) than a trade on the spot market today.

10. The process of protecting oneself from the riskiness of exchange rate movements; one method is by entering into a forward contract.

When the forward ER is such that a forward trade costs more than a spot trade today costs, there is said to be a **forward premium**⁸. If the reverse were true, such that the forward trade were cheaper than a spot trade, then there is a **forward discount**⁹.

A currency trader is **hedging**¹⁰ if he or she enters into a forward contract to protect oneself from a downside loss. However, by hedging the trader also forfeits the potential for an upside gain. Suppose in the story above that the spot ER falls rather

than rises. Suppose the ER fell to 1.10 \$/€. In this case, had the importer waited, the €1,000,000 would only have cost \$1,100,000 (i.e., $\$1,000,000 \times 1.10 \text{ \$/\euro}$). Thus hedging protects against loss but at the same time eliminates potential unexpected gain.

KEY TAKEAWAYS

- An exchange rate denominated x/y gives the value of y in terms of x . When an exchange rate denominated x/y rises, then y has appreciated in value in terms of x , while x has depreciated in terms of y .
- Spot exchange rates represent the exchange rate prevailing for currency trades today. Forward, or future, exchange rates represent the exchange values on trades that will take place in the future to fulfill a predetermined contract.
- Currency arbitrage occurs when someone buys a currency at a low price and sells shortly afterward at a higher price to make a profit.
- Hedging refers to actions taken to reduce the risk associated with currency trades.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe an increase in the value of the yen.
 - b. This currency value is expressed by the euro/peso exchange rate.
 - c. This has happened to the value of the U.S. dollar if the dollar/euro exchange rate rises from 1.10 \$/€ to 1.20 \$/€.
 - d. The term used to describe the process of buying low and selling high to make a profit.
 - e. The term used to describe the exchange rate that appears on a contract to exchange currencies either 30, 60, 90, or 180 days in the future.
 - f. The term used to describe the exchange rate that prevails for (almost) immediate trades.
 - g. The term used to describe process of protecting oneself from the riskiness of exchange rate movements.
2. Use the exchange rate data in the table to answer the following questions. The first two exchange rates are the spot rates on those dates. The third exchange rate is the one-year forward exchange rate as of February 2004.

	February 4, 2003	February 4, 2004	Forward February 4, 2005
United States–Europe	1.08 \$/€	1.25 \$/€	1.24 \$/€
South Africa–United States	8.55 rand/\$	6.95 rand/\$	7.42 rand/\$

- a. Calculate the rate of change in the euro value relative to the dollar between 2003 and 2004.
- b. Calculate the rate of change in the dollar value relative to the euro between 2003 and 2004.

- c. Calculate the rate of change in the dollar value relative to the South African rand between 2003 and 2004.
- d. Calculate the expected change in the dollar value relative to the euro between 2004 and 2005.
- e. Calculate the expected change in the dollar value relative to the rand between 2004 and 2005.

15.3 Calculating Rate of Returns on International Investments

LEARNING OBJECTIVE

1. Learn how to calculate the rate of return (RoR) for a domestic deposit and a foreign deposit.

Suppose that an investor holding U.S. dollars must decide between two investments of equal risk and liquidity. Suppose one potential investment is a one-year **certificate of deposit (CD)**¹¹ issued by a U.S. bank while a second potential investment is a one-year CD issued by a British bank. For simplicity we'll assume that interest is calculated on both CDs using a simple interest rather than with a compounding formula. A CD is a type of deposit that provides a higher rate of interest to the depositor in return for a promise to keep the money deposited for a fixed amount of time. The time period could be six months, one year, two years, or any other period decided by the bank. If the depositor wants to withdraw the money earlier, she must pay a penalty.

Since we imagine that an investor wants to obtain the highest rate of return (RoR) possible, given acceptable risk and liquidity characteristics, that investor will choose the investment with the highest rate of return. If the investor acted naively, she might simply compare interest rates between the two investments and choose the one that is higher. However, this would not necessarily be the best choice. To see why, we need to walk through the calculation of rates of return on these two investments.

First, we need to collect some data, which we will do in general terms rather than use specific values. Examples with actual values are presented in a later section.

Let $E_{\$/\text{£}}$ = the spot ER.

$E_{\$/\text{£}}^e$ = the expected ER one year from now.

$i_{\$}$ = the one-year interest rate on a CD in the United States (in decimal form).

$i_{\text{£}}$ = the one-year interest rate on a CD in Britain (in decimal form).

11. A type of deposit that provides a higher rate of interest to the depositor in return for a promise to keep the money deposited for a fixed amount of time.

U.S. Rate of Return

The rate of return on the U.S. CD is simply the interest rate on that deposit. More formally,

$$RoR_{\$} = i_{\$}.$$

This is because the interest rate describes the percentage increase in the value of the deposit over the course of the year. It is also simple because there is no need to convert currencies.

British Rate of Return

The rate of return on the British CD is more difficult to determine. If a U.S. investor, with dollars, wants to invest in the British CD, she must first exchange dollars for pounds on the spot market and then use the British pound (£) to purchase the British CD. After one year, she must convert pounds back to dollars at the exchange rate that prevails then. The rate of return on that investment is the percentage change in dollar value during the year. To calculate this we can follow the procedure below.

Suppose the investor has P dollars to invest (P for principal).

Step 1: Convert the dollars to pounds.

$\frac{P}{E_{\$/\text{\textsterling}}}$ is the number of pounds the investor will have at the beginning of the year.

Step 2: Purchase the British CD and earn interest in pounds during the year.

$\frac{P}{E_{\$/\text{\textsterling}}} (1 + i_{\text{\textsterling}})$ is the number of pounds the investor will have at the end of the year. The first term in parentheses returns the principal. The second term is the interest payment.

Step 3: Convert the principal plus interest back into dollars in one year.

$\frac{P}{E_{\$/\text{\textsterling}}} (1 + i_{\text{\textsterling}}) E_{\$/\text{\textsterling}}^e$ is the number of dollars the investor can expect to have at the end of the year.

The rate of return in dollar terms from this British investment can be found by calculating the expected percentage change in the value of the investor's dollar assets over the year, as shown below:

$$RoR_{\text{\textsterling}} = \frac{\frac{P}{E_{\$/\text{\textsterling}}} (1 + i_{\text{\textsterling}}) E_{\$/\text{\textsterling}}^e - P}{P}.$$

After factoring out the P , this reduces to

$$RoR_{\text{\textsterling}} = \frac{E_{\$/\text{\textsterling}}^e}{E_{\$/\text{\textsterling}}} (1 + i_{\text{\textsterling}}) - 1.$$

Thus the rate of return on the foreign investment is more complicated because the set of transactions is more complicated. For the U.S. investment, the depositor simply deposits the dollars and earns dollar interest at the rate given by the interest rate. However, for the foreign deposit, the investor must first convert currency, then deposit the money abroad earning interest in foreign currency units, and finally reconvert the currency back to dollars. The rate of return depends not only on the foreign interest rate but also on the spot exchange rate and the expected exchange rate one year in the future.

Note that according to the formula, the rate of return on the foreign deposit is positively related to changes in the foreign interest rate and the expected foreign currency value and negatively related to the spot foreign currency value.

KEY TAKEAWAYS

- For a dollar investor, the rate of return on a U.S. deposit is equal to the interest rate: $RoR\$ = i\$$.
- For a dollar investor, the rate of return on a foreign deposit depends on the foreign interest rate, the spot exchange rate, and the exchange rate expected to prevail at the time the deposit is redeemed: In particular,
$$RoR_f = \frac{E_{\$/f}^e}{E_{\$/f}} (1 + i_f) - 1.$$

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. These three variables influence the rate of return on a foreign deposit.
 - b. For a U.S. dollar investor, this is the rate of return on a U.S. dollar deposit yielding 3 percent per year.
 - c. The term used to describe the exchange rate predicted to prevail at some point in the future.
 - d. The term for the type of bank deposit that offers a higher yield on a deposit that is maintained for a predetermined period of time.

15.4 Interpretation of the Rate of Return Formula

LEARNING OBJECTIVE

- Break down the rate of return on foreign deposits into three distinct components.

Although the derivation of the rate of return formula is fairly straightforward, it does not lend itself easily to interpretation or intuition. By applying some algebraic “tricks,” it is possible to rewrite the British rate of return formula in a form that is much more intuitive.

Step 1: Begin with the British rate of return formula derived in [Chapter 15 "Foreign Exchange Markets and Rates of Return", Section 15.3 "Calculating Rate of Returns on International Investments"](#):

$$RoR_{\text{£}} = \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} (1 + i_{\text{£}}) - 1.$$

Step 2: Factor out the term in parentheses. Add $i_{\text{£}}$ and then subtract it as well. Mathematically, a term does not change in value if you add and subtract the same value:

$$RoR_{\text{£}} = \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} + i_{\text{£}} \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} - 1 + i_{\text{£}} - i_{\text{£}}.$$

Step 3: Change the (-1) in the expression to its equivalent, $-\frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}}$. Also change $-i_{\text{£}}$ to its equivalent, $-i_{\text{£}} \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}}$. Since $\frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} = 1$, these changes do not change the value of the rate of return expression:

$$RoR_{\text{£}} = \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} + i_{\text{£}} \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} - \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} + i_{\text{£}} - i_{\text{£}} \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}}.$$

Step 4: Rearrange the expression:

$$RoR_{\text{£}} = i_{\text{£}} + \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}} + i_{\text{£}} \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}} - i_{\text{£}} \frac{E_{\$/\text{£}}}{E_{\$/\text{£}}}.$$

Step 5: Simplify by combining terms with common denominators:

$$RoR_{\text{£}} = i_{\text{£}} + \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}} + i_{\text{£}} \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}.$$

Step 6: Factor out the percentage change in the exchange rate term:

$$RoR_{\text{£}} = i_{\text{£}} + (1 + i_{\text{£}}) \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}.$$

This formula shows that the expected rate of return on the British asset depends on two things, the British interest rate and the expected percentage change in the value of the pound. Notice that if $\frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}$ is a positive number, then the expected $\$/\text{£}$ ER is greater than the current spot ER, which means that one expects a pound appreciation in the future. Furthermore, $\frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}$ represents the expected rate of appreciation of the pound during the following year. Similarly, if $\frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}$ were negative, then it corresponds to the expected rate of depreciation of the pound during the subsequent year.

The expected rate of change in the pound value is multiplied by $(1 + i_{\text{£}})$, which generally corresponds to a principal and interest component in a rate of return calculation.

To make sense of this expression, it is useful to consider a series of simple numerical examples.

Suppose the following values prevail,

$i_{\text{£}}$	5% per year
$E_{\$/\text{£}}^{\text{e}}$	1.1 $\$/\text{£}$
$E_{\$/\text{£}}$	1.0 $\$/\text{£}$

Plugging these into the rate of return formula yields

$$RoR_{\text{£}} = 0.05 + (1 + 0.05) \frac{1.10 - 1.00}{1.00}, \infty$$

which simplifies to

$$RoR_{\text{£}} = 0.05 + (1 + 0.05) \times 0.10 = .155 \text{ or } 15.5\%.$$

Note that because of the exchange rate change, the rate of return on the British asset is considerably higher than the 5 percent interest rate.

To decompose these effects suppose that the British asset yielded no interest whatsoever.

This would occur if the individual held pound currency for the year rather than purchasing a CD. In this case, the rate of return formula reduces to

$$RoR_{\text{£}} = 0.0 + (1 + 0.0) \times 0.10 = .10 \text{ or } 10\%.$$

This means that 10 percent of the rate of return arises solely because of the pound appreciation. Essentially an investor in this case gains because of currency arbitrage over time. Remember that arbitrage means buying something when its price is low, selling it when its price is high, and thus making a profit on the series of transactions. In this case, the investor buys pounds at the start of the year, when their price (in terms of dollars) is low, and then resells them at the end of the year when their price is higher.

Next, suppose that there was no exchange rate change during the year, but there was a 5 percent interest rate on the British asset. In this case, the rate of return becomes

$$RoR_{\text{£}} = 0.05 + (1 + 0.05) \times 0.0 = .05 \text{ or } 5\%.$$

Thus with no change in the exchange rate, the rate of return reduces to the interest rate on the asset.

Finally, let's look back at the rate of return formula:

$$RoR_{\text{£}} = i_{\text{£}} + (1 + i_{\text{£}}) \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}.$$

The first term simply gives the contribution to the total rate of return that derives solely from the interest rate on the foreign asset. The second set of terms has the percentage change in the exchange rate times one plus the interest rate. It corresponds to the contribution to the rate of return that arises solely due to the exchange rate change. The one plus interest rate term means that the exchange rate return can be separated into two components, a principal component and an interest component.

Suppose the exchange rate change is positive. In this case, the principal that is originally deposited will grow in value by the percentage exchange rate change. But the principal also accrues interest and as the £ value rises, the interest value, in dollar terms, also rises.

Thus the second set of terms represents the percentage increase in the value of one's principal and interest that arises solely from the change in the exchange rate.

KEY TAKEAWAYS

- The rate of return on a foreign deposit consists of three components: the interest rate itself, the change in the value of the principal due to the exchange rate change, and the change in the value of the interest due to the exchange rate change.
- Another formula, but one that is equivalent to the one in the previous section, for the rate of return on a foreign deposit is

$$RoR_{\text{£}} = i_{\text{£}} + (1 + i_{\text{£}}) \frac{E_{\$/\text{£}}^e - E_{\$/\text{£}}}{E_{\$/\text{£}}}.$$

EXERCISES

1. Consider the following data. Suppose the expected exchange rates are the average expectations by investors for exchange rates in one year. Imagine that the interest rates are for equally risky assets and are annual rates.

	United States	Australia	Singapore
Current Exchange Rate	-	1.80 A\$/US\$	1.75 S\$/US\$
Expected Exchange Rate	-	1.90 A\$/US\$	1.65 S\$/US\$
Current Interest Rate (%)	2.0	4.0	1.0

- a. Calculate the rate of return for a U.S. dollar investor investing in the Australian deposit for one year.
 - b. Calculate the rate of return for a U.S. dollar investor investing in the Singapore deposit for one year.
 - c. Among these three options (United States, Australia, and Singapore), which is the best place for the investor to invest? Which is the worst place?
2. The covered interest parity condition substitutes the forward exchange rate for the expected exchange rate. The condition is labeled “covered” because the forward contract assures a certain rate of return (i.e., without risk) on foreign deposits. The table below lists a spot exchange rate, a ninety-day forward rate, and a ninety-day money market interest rate in Germany and Canada. Use this information to answer the following questions.

	Germany	Canada
Spot Exchange Rate	0.5841 \$/DM	0.7451 US\$/C\$
90-Day Forward Exchange Rate	0.5807 \$/DM	0.7446 US\$/C\$
90-Day Interest Rate (%)	1.442	0.875

What would the U.S. ninety-day interest rate have to be for the United States to have the highest rate of return for a U.S. investor? (Use the exact formulas to calculate the rates of return.)

15.5 Applying the Rate of Return Formulas

LEARNING OBJECTIVE

1. Learn how to apply numerical values for exchange rates and interest rates to the rate of return formulas to determine the best international investment.

Use the data in the tables below to calculate in which country it would have been best to purchase a one-year interest-bearing asset. These numbers were taken from the *Economist*, Weekly Indicators, December 17, 2005, p. 90, <http://www.economist.com>.

Example 1

Consider the following data for interest rates and exchange rates in the United States and Britain:

$i_{\$}$	2.37% per year
$i_{\text{£}}$	4.83% per year
$E_{\$/\text{£}}^{04}$	1.96 \\$/£
$E_{\$/\text{£}}^{05}$	1.75 \\$/£

We imagine that the decision is to be made in 2004, looking forward into 2005. However, we calculate this in hindsight after we know what the 2005 exchange rate is. Thus we plug in the 2005 rate for the expected exchange rate and use the 2004 rate as the current spot rate. Thus the ex-post (i.e., after the fact) rate of return on British deposits is given by

$$RoR_{\text{£}} = 0.0483 + (1 + 0.0483) \frac{1.75 - 1.96}{1.96},$$

which simplifies to

$$RoR_{\text{£}} = 0.0483 + (1 + 0.0483)(-0.1071) = -0.064 \text{ or } -6.4\%.$$

A negative rate of return means that the investor would have lost money (in dollar terms) by purchasing the British asset.

Since $\text{RoR}_\$ = 2.37\% > \text{RoR}_\£ = -6.4\%$, the investor seeking the highest rate of return should have deposited her money in the U.S. account.

Example 2

Consider the following data for interest rates and exchange rates in the United States and Japan.

$i_\$$	2.37% per year
$i_\text{¥}$	0.02% per year
$E_{\text{¥}/\04	104 ¥/\$
$E_{\text{¥}/\05	120 ¥/\$

Again, imagine that the decision is to be made in 2004, looking forward into 2005. However, we calculate this in hindsight after we know what the 2005 exchange is. Thus we plug in the 2005 rate for the expected exchange rate and use the 2004 rate as the current spot rate. Note also that the interest rate in Japan *really* was 0.02 percent. It was virtually zero.

Before calculating the rate of return, it is necessary to convert the exchange rate to the yen equivalent rather than the dollar equivalent. Thus

$$E_{\$/\text{¥}}^{04} = \frac{1}{104} = 0.0096 \text{ and } E_{\$/\text{¥}}^{05} = \frac{1}{120} = 0.0083.$$

Now, the ex-post (i.e., after the fact) rate of return on Japanese deposits is given by

$$\text{RoR}_\text{¥} = 0.0002 + (1 + 0.0002) \frac{0.0083 - 0.0096}{0.0096},$$

which simplifies to

$$\text{RoR}_\text{¥} - 0.0002 + (1 + 0.0002)(-0.1354) = -0.1352 \text{ or } -13.52\%.$$

A negative rate of return means that the investor would have lost money (in dollar terms) by purchasing the Japanese asset.

Since $\text{RoR}_\$ = 2.37\% > \text{RoR}_¥ = -13.52\%$, the investor seeking the highest rate of return should have deposited his money in the U.S. account.

Example 3

Consider the following data for interest rates and exchange rates in the United States and South Korea. Note that South Korean currency is in won (W).

$i_{\$}$	2.37% per year
i_W	4.04% per year
$E_{W/\04	1,059 W/\$
$E_{W/\05	1,026 W/\$

As in the preceding examples, the decision is to be made in 2004, looking forward to 2005. However, since the previous year interest rate is not listed, we use the current short-term interest rate. Before calculating the rate of return, it is necessary to convert the exchange rate to the won equivalent rather than the dollar equivalent. Thus

$$E_{\$/W}^{04} = \frac{1}{1059} = 0.000944 \text{ and } E_{\$/W}^{05} = \frac{1}{1026} = 0.000975.$$

Now, the ex-post (i.e., after the fact) rate of return on Italian deposits is given by

$$\text{RoR}_W = 0.0404 + (1 + 0.0404) \frac{0.000975 - 0.000944}{0.000944},$$

which simplifies to

$$\text{RoR}_W = 0.0404 + (1 + 0.0404)(0.0328) = 0.0746 \text{ or } +7.46\%.$$

In this case, the positive rate of return means an investor would have made money (in dollar terms) by purchasing the South Korean asset.

Also, since $RoR\$ = 2.37$ percent < $RoRW = 7.46$ percent, the investor seeking the highest rate of return should have deposited his money in the South Korean account.

KEY TAKEAWAY

- An investor should choose the deposit or asset that promises the highest expected rate of return assuming equivalent risk and liquidity characteristics.

EXERCISES

1. Consider the following data collected on February 9, 2004. The interest rate given is for a one-year money market deposit. The spot exchange rate is the rate for February 9. The expected exchange rate is the one-year forward rate. Express each answer as a percentage.

$i_{C\$}$	2.5%
$E_{US\$/C\$}$	0.7541 US\\$/C\$
$E_{US\$/C\e	0[0].7468 US\\$/C\$

- a. Use both RoR formulas (one from [Chapter 15 "Foreign Exchange Markets and Rates of Return", Section 15.3 "Calculating Rate of Returns on International Investments"](#), the other from [Chapter 15 "Foreign Exchange Markets and Rates of Return", Section 15.4 "Interpretation of the Rate of Return Formula"](#), Step 5) to calculate the expected rate of return on the Canadian money market deposit and show that both formulas generate the same answer.
 - b. What part of the rate of return arises only due to the interest earned on the deposit?
 - c. What part of the rate of return arises from the percentage change in the value of the principal due to the change in the exchange rate?
 - d. What component of the rate of return arises from the percentage change in the value of the interest payments due to the change in the exchange rate?
2. Consider the following data collected on February 9, 2004. The interest rate given is for a one-year money market deposit. The spot exchange rate is the rate for February 9. The expected exchange rate is the one-year forward rate. Express each answer as a percentage.

$i_{\text{£}}$	4.5%
$E_{\$/\text{£}}$	1.8574 \\$/£
$E_{\$/\text{£}}^e$	1.7956 \\$/£

- a. Use both RoR formulas (one from Chapter 15 "Foreign Exchange Markets and Rates of Return", Section 15.3 "Calculating Rate of Returns on International Investments", the other from Chapter 15 "Foreign Exchange Markets and Rates of Return", Section 15.4 "Interpretation of the Rate of Return Formula", Step 5) to calculate the expected rate of return on the British money market deposit and show that both formulas generate the same answer.
- b. What part of the rate of return arises only due to the interest earned on the deposit?
- c. What part of the rate of return arises from the percentage change in the value of the principal due to the change in the exchange rate?
- d. What component of the rate of return arises from the percentage change in the value of the interest payments due to the change in the exchange rate?

Chapter 16

Interest Rate Parity

Interest rate parity is one of the most important theories in international finance because it is probably the best way to explain how exchange rate values are determined and why they fluctuate as they do. Most of the international currency exchanges occur for investment purposes, and therefore understanding the prime motivations for international investment is critical.

The chapter applies the rate of return formula developed in [Chapter 15 "Foreign Exchange Markets and Rates of Return"](#) and shows how changes in the determinants of the rate of return on assets affect investor behavior on the foreign exchange market, which in turn affects the value of the exchange rate. The model is described in two different ways: first, using simple supply and demand curves; and second, using a rate of return diagram that will be used later with the development of a more elaborate macro model of the economy.

16.1 Overview of Interest Rate Parity

LEARNING OBJECTIVES

1. Define the interest rate parity condition.
2. Learn the asset approach to exchange rate determination.

Interest rate parity (IRP)¹ is a theory used to explain the value and movements of exchange rates. It is also known as the asset approach to exchange rate determination. The **interest rate parity theory**² assumes that the actions of international investors—motivated by cross-country differences in rates of return on comparable assets—induce changes in the spot exchange rate. In another vein, IRP suggests that transactions on a country's financial account affect the value of the exchange rate on the foreign exchange (Forex) market. This contrasts with the purchasing power parity theory, which assumes that the actions of importers and exporters, whose transactions are recorded on the current account, induce changes in the exchange rate.

Interest Rate Parity Condition

Interest rate parity refers to a condition of equality between the rates of return on comparable assets between two countries. The term is somewhat of a misnomer on the basis of how it is being described here, as it should really be called rate of return parity. The term developed in an era when the world was in a system of fixed exchange rates. Under those circumstances, and as will be demonstrated in a later chapter, rate of return parity did mean the equalization of interest rates. However, when exchange rates can fluctuate, interest rate parity becomes rate of return parity, but the name was never changed.

In terms of the rates of return formulas developed in [Chapter 15 "Foreign Exchange Markets and Rates of Return"](#), interest rate parity holds when the rate of return on dollar deposits is just equal to the expected rate of return on British deposits, that is, when

$$RoR_{\$} = RoR_E.$$

Plugging in the above formula yields

1. A condition in which the rates of return on comparable assets in two countries are equal.
2. A theory of exchange rate determination based on investor motivations in which equilibrium is described by the interest rate parity condition.

$$i_{\$} = i_{\text{£}} + (1 + i_{\text{£}}) \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}.$$

This condition is often simplified in many textbooks by dropping the final term in which the British interest rate is multiplied by the exchange rate change. The logic is that the final term is usually very small especially when interest rates are low. The *approximate version* of the IRP condition then is

$$i_{\$} - i_{\text{£}} = \frac{E_{\$/\text{£}}^{\text{e}} - E_{\$/\text{£}}}{E_{\$/\text{£}}}.$$

One should be careful, however. The approximate version would not be a good approximation when interest rates in a country are high. For example, back in 1997, short-term interest rates were 60 percent per year in Russia and 75 percent per year in Turkey. With these interest rates, the approximate formula would not give an accurate representation of rates of return.

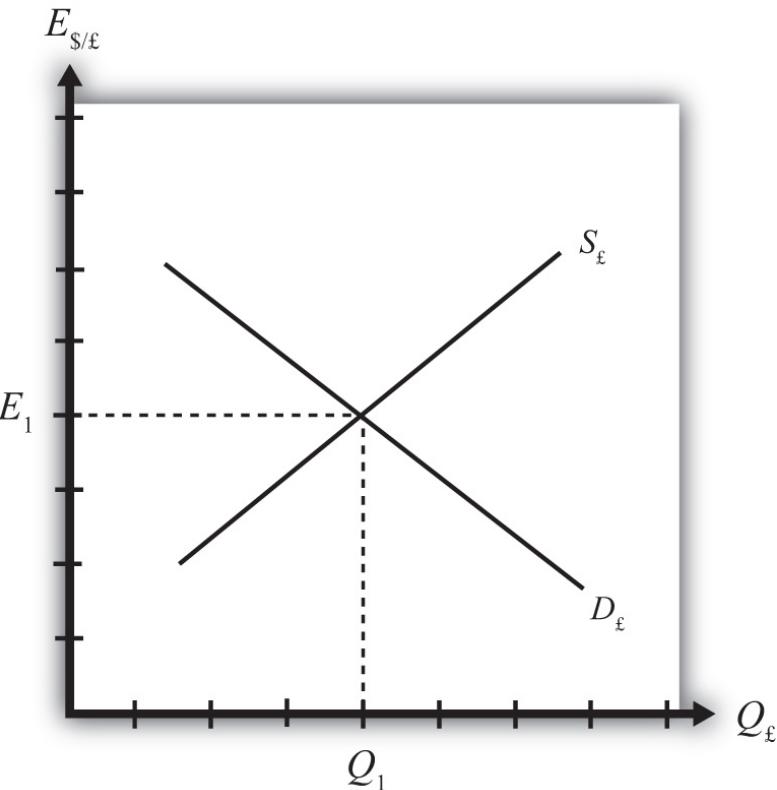
Interest Rate Parity Theory

Investor behavior in asset markets that results in interest parity can also explain why the exchange rate may rise and fall in response to market changes. In other words, interest parity can be used to develop a model of exchange rate determination. This is known as the asset approach, or the interest rate parity model.

The first step is to reinterpret the rate of return calculations described previously in more general (aggregate) terms. Thus instead of using the interest rate on a one-year certificate of deposit (CD), we will interpret the interest rates in the two countries as the average interest rates that currently prevail. Similarly, we will imagine that the expected exchange rate is the average expectation across many different individual investors. The rates of return then are the average expected rates of return on a wide variety of assets between two countries.

Next, we imagine that investors trade currencies in the foreign exchange (Forex) market. Each day, some investors come to a market ready to supply a currency in exchange for another, while others come to demand currency in exchange for another.

Figure 16.1 The Forex for British Pounds



Consider the market for British pounds (£) in New York depicted in [Figure 16.1 "The Forex for British Pounds"](#). We measure the supply and demand of pounds along the horizontal axis and the price of pounds (i.e., the exchange rate $E_{\$/\text{£}}$) on the vertical axis. Let $S_\text{£}$ represent the supply of pounds in exchange for dollars at all different exchange rates that might prevail. The supply is generally by British investors who demand dollars to purchase dollar denominated assets. However, supply of pounds might also come from U.S. investors who decide to convert previously acquired pound currency. Let $D_\text{£}$ the demand for pounds in exchange for dollars at all different exchange rates that might prevail. The demand is generally by U.S. investors who supply dollars to purchase pound-denominated assets. Of course, demand may also come from British investors who decide to convert previously purchased dollars. Recall that

$$RoR_\text{£} = i_\text{£} + (1 + i_\text{£}) \frac{E_{\$/\text{£}}^\text{e} - E_{\$/\text{£}}}{E_{\$/\text{£}}},$$

which implies that as $E_{\$/\text{£}}$ rises, $RoR_\text{£}$ falls. This means that British investors would seek to supply more pounds at higher pound values but U.S. investors would

demand fewer pounds at higher pound values. This explains why the supply curve slopes upward and the demand curve slopes downward.

The intersection of supply and demand specifies the equilibrium exchange rate (E_1) and the quantity of pounds (Q_1) traded in the market. When the Forex is at equilibrium, it must be that interest rate parity is satisfied. This is true because the violation of interest rate parity will cause investors to shift funds from one country to another, thereby causing a change in the exchange rate. This process is described in more detail in [Chapter 16 "Interest Rate Parity"](#), [Section 16.2 "Comparative Statics in the IRP Theory"](#).

KEY TAKEAWAYS

- Interest rate parity in a floating exchange system means the equalization of rates of return on comparable assets between two different countries.
- Interest rate parity is satisfied when the foreign exchange market is in equilibrium, or in other words, IRP holds when the supply of currency is equal to the demand in the Forex.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This theory of exchange rate determination is also known as the asset approach.
 - b. The name of the condition in which rates of return on comparable assets in different countries are equal.
 - c. Of *greater, less, or equal*, this is how the supply of pounds compares to the demand for pounds in the foreign exchange market when interest rate parity holds.

16.2 Comparative Statics in the IRP Theory

LEARNING OBJECTIVE

1. Learn how changes in interest rates and expected exchange rates can influence international investment decisions and affect the exchange rate value.

Comparative statics refers to an exercise in a model that assesses how changes in an exogenous variable will affect the values of the endogenous variables. The endogenous variables are those whose values are determined in the equilibrium. In the IRP model, the endogenous variables are the exchange rate value and—of lesser importance—the quantity of currencies exchanged on the Forex market. The exogenous variables are those whose values are given beforehand and are known by the model's decision makers. In the IRP model, the exogenous variables are those that influence the positions of the rate of return curves, including the U.S. interest rate, the British interest rate, and the expected future exchange rate. Another way to describe this is that the endogenous variable values are determined within the model, while the exogenous variable values are determined outside of the model.

Comparative statics exercises enable one to answer a question like “What would happen to the exchange rate if there were an increase in U.S. interest rates?” When assessing a question like this, economists will invariably invoke the **ceteris paribus**³ assumption. Ceteris paribus means that we assume all other exogenous variables are maintained at their original values when we change the variable of interest. Thus if we assess what would happen to the exchange rate (an endogenous variable) if there were an increase in the U.S. interest rate (an exogenous variable) while invoking ceteris paribus, then ceteris paribus means keeping the original values for the other exogenous variables (in this case, the British interest rate and the expected future exchange rate) fixed.

It is useful to think of a comparative statics exercise as a controlled economic experiment. In the sciences, one can test propositions by controlling the environment of a physical system in such a way that one can isolate the particular cause-and-effect relationship. Thus, to test whether a ball and a feather will fall at the same rate in a frictionless vacuum, experimenters could create a vacuum environment and measure the rate of descent of the ball versus the feather. In economic systems, such experiments are virtually impossible because one can never eliminate all the “frictions.”

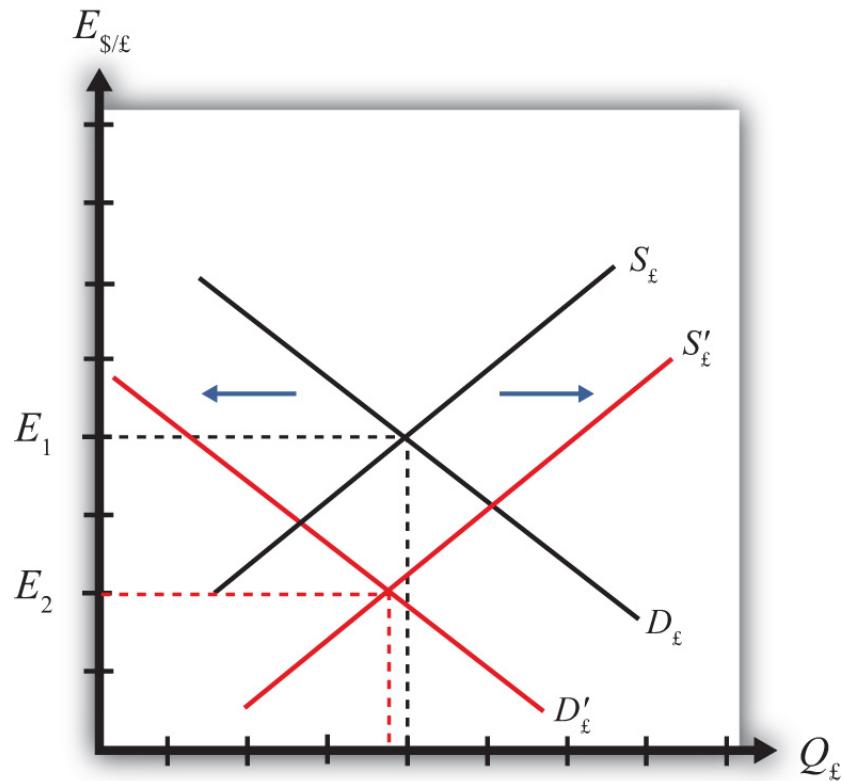
3. Means that all other exogenous variables are maintained at their original values when a change in one exogenous variable takes place.

However, by creating mathematical economic systems (i.e., an economic model), it becomes possible to conduct similar types of “experiments.” A comparative statics exercise allows one to isolate how a change in one exogenous variable affects the value of the equilibrium variable while controlling for changes in other variables that might also affect the outcome.

The Effect of Changes in U.S. Interest Rates on the Spot Exchange Rate

Suppose that the Forex is initially in equilibrium such that $S_E = D_E$ at the exchange rate E_1 . Now let average U.S. interest rates ($i_{\$}$) rise, ceteris paribus. The increase in interest rates raises the rate of return on U.S. assets ($RoR_{\$}$), which at the original exchange rate causes the rate of return on U.S. assets to exceed the rate of return on British assets ($RoR_{\$} > RoR_E$). This will raise the supply of pounds on the Forex as British investors seek the higher average return on U.S. assets. It will also lower the demand for British pounds (£) by U.S. investors who decide to invest at home rather than abroad.

Figure 16.2 Effects of a U.S. Interest Rate Increase

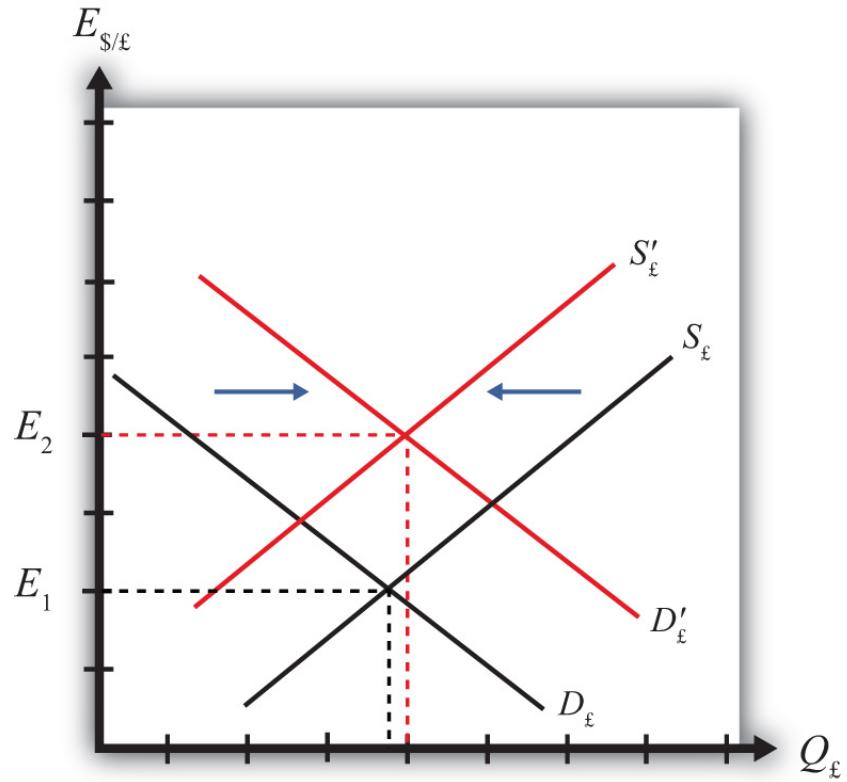


Thus in terms of the Forex market depicted in [Figure 16.2 "Effects of a U.S. Interest Rate Increase"](#), S_E shifts right (black to red) while D_E shifts left (black to red). The equilibrium exchange rate falls to E_2 . This means that the increase in U.S. interest rates causes a pound depreciation and a dollar appreciation. As the exchange rate falls, RoR_E rises since $RoR_E = \frac{E_{\$/\text{£}}}{E_{\$/\text{£}}^e} (1 + i_{\text{£}}) - 1$. RoR_E continues to rise until the interest parity condition, $RoR\$ = RoR_E$, again holds.

The Effect of Changes in British Interest Rates on the Spot Exchange Rate

Suppose that the Forex is initially in equilibrium such that $S_E = D_E$ at the exchange rate E_1 shown in [Figure 16.3 "Effects of a British Interest Rate Increase"](#). Now let average British interest rates (i_E) rise, ceteris paribus. The increase in interest rates raises the rate of return on British assets (RoR_E), which at the original exchange rate causes the rate of return on British assets to exceed the rate of return on U.S. assets ($RoR_E > RoR\$$).

Figure 16.3 Effects of a British Interest Rate Increase



This will raise the demand for pounds on the Forex as U.S. investors seek the higher average return on British assets. It will also lower the supply of British pounds by British investors who decide to invest at home rather than abroad. Thus in terms of the graph, D_E shifts right (black to red) while S_E shifts left (black to red). The equilibrium exchange rate rises to E_2 . This means that the increase in British interest rates causes a pound appreciation and a dollar depreciation. As the exchange rate rises, RoR_E falls since $RoR_E = \frac{E_{\$/E}^e}{E_{\$/E}} (1 + i_E) - 1$. RoR_E continues to fall until the interest parity condition, $RoR\$ = RoR_E$, again holds.

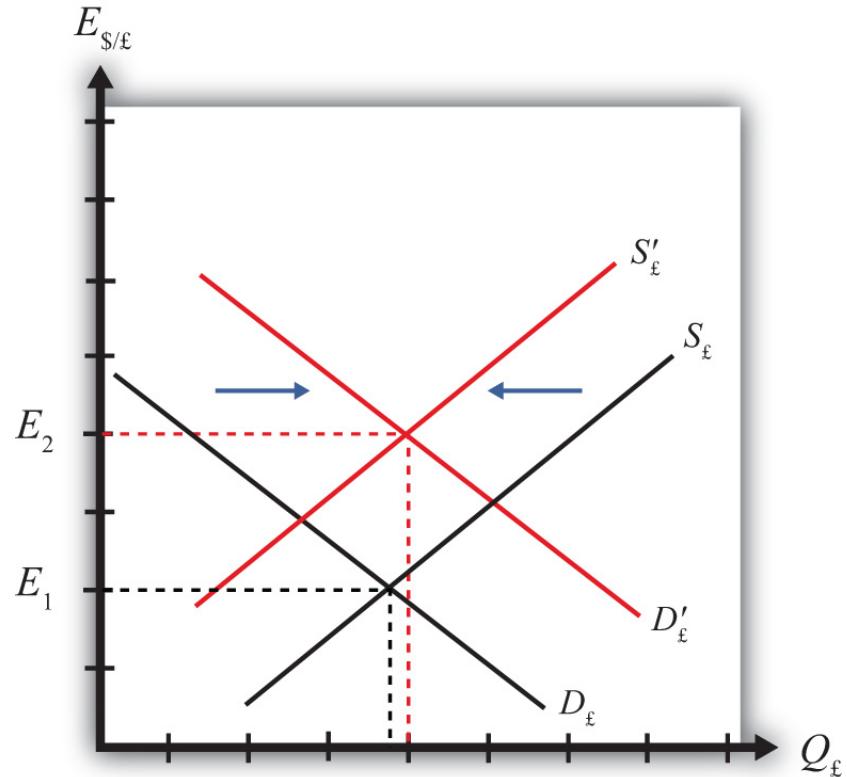
The Effect of Changes in the Expected Exchange Rate on the Spot Exchange Rate

Suppose that the Forex is initially in equilibrium such that $S_E = D_E$ at the exchange rate E_1 . Now suppose investors suddenly raise their expected future exchange rate ($E_{\$/E}^e$), ceteris paribus. This means that if investors had expected the pound to appreciate, they now expect it to appreciate more. Likewise, if investors had expected the dollar to depreciate, they now expect it to depreciate more. Also, if they had expected the pound to depreciate, they now expect it to depreciate less. Likewise, if they had expected the dollar to appreciate, they now expect it to appreciate less.

This change might occur because new information is released. For example, the British Central Bank might release information that suggests an increased chance that the pound will rise in value in the future.

The increase in the expected exchange rate raises the rate of return on British assets (RoR_E), which at the original exchange rate causes the rate of return on British assets to exceed the rate of return on U.S. assets ($RoR_E > RoR\$$). This will raise the demand for the pound on the Forex as U.S. investors seek the higher average return on British assets. It will also lower the supply of British pounds by British investors who decide to invest at home rather than abroad. Thus, as depicted in Figure 16.4 "Effects of a Change in the Expected Exchange Rate", D_E shifts right (black to red) while S_E shifts left (black to red). The equilibrium exchange rate rises to E_2 . This means that the increase in the expected exchange rate ($E_{\$/E}^e$) causes a pound appreciation and a dollar depreciation.

Figure 16.4 Effects of a Change in the Expected Exchange Rate



This is a case of self-fulfilling expectations. If investors suddenly think the pound will appreciate more in the future and if they act on that belief, then the pound will begin to rise in the present, hence fulfilling their expectations.

As the exchange rate rises, $RoR_{\mathcal{L}}$ falls since $RoR_{\mathcal{L}} = \frac{E_{\$/\mathcal{L}}^e}{E_{\$/\mathcal{L}}} (1 + i_{\mathcal{L}}) - 1$. $RoR_{\mathcal{L}}$ continues to fall until the interest parity condition, $RoR_{\$} = RoR_{\mathcal{L}}$, again holds.

KEY TAKEAWAYS

- An increase in U.S. interest rates causes a pound depreciation and a dollar appreciation.
- An increase in British interest rates causes a pound appreciation and a dollar depreciation.
- An increase in the expected exchange rate ($E_{\$/\mathcal{L}}^e$) causes a pound appreciation and a dollar depreciation.

EXERCISES

1. Consider the economic changes listed along the left column of the following table. Indicate the effect of each change on the variables listed in the first row. Use insights from the interest rate parity model to determine the answers. Assume floating exchange rates. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	U.S. Dollar Value	$E\$/\epsilon$
a. A decrease in U.S. interest rates		
b. An increase in expected U.S. economic growth that raises expected asset values		
c. An expected increase in European stock values		

2. On February 5, 2004, the *Wall Street Journal* reported that Asian central banks were considering selling a significant share of their U.S. government bond holdings. It was estimated at the time that foreign central banks owned over \$800 billion in U.S. Treasury bonds, or one-fifth of all U.S. federal government debt. Taiwan was considering using some of its foreign reserves to help its businesses purchase U.S. machinery.
- a. What is the likely effect on the U.S. dollar value if Taiwan implements its plan? Explain.

- b. What effect would this transaction have on the U.S. trade deficit? Explain.
- c. How would the answer to part a change if the Taiwanese government used sales of its foreign reserves to help its businesses purchase Taiwanese-produced machinery? Explain.

16.3 Forex Equilibrium with the Rate of Return Diagram

LEARNING OBJECTIVE

1. Use the rate of return plots to represent the interest rate parity equilibrium in the foreign exchange market.

An alternative graphical approach is sometimes used to depict the equilibrium exchange rate in the foreign exchange (Forex) market. The graph is called the rate of return diagram since it depicts rates of return for assets in two separate countries as functions of the exchange rate. The equilibrium condition depicted in the diagram represents the interest rate parity condition. In effect, the diagram identifies the equilibrium exchange rate that must prevail to satisfy the interest rate parity condition.

Recall the rate of return formulas for deposits in two separate countries. Consider an investor, holding U.S. dollars, comparing the purchase of a one-year certificate of deposit (CD) at a U.S. bank with a one-year CD issued by a British bank. The rate of return on the U.S. deposit works out simply to be the U.S. interest rate shown below:

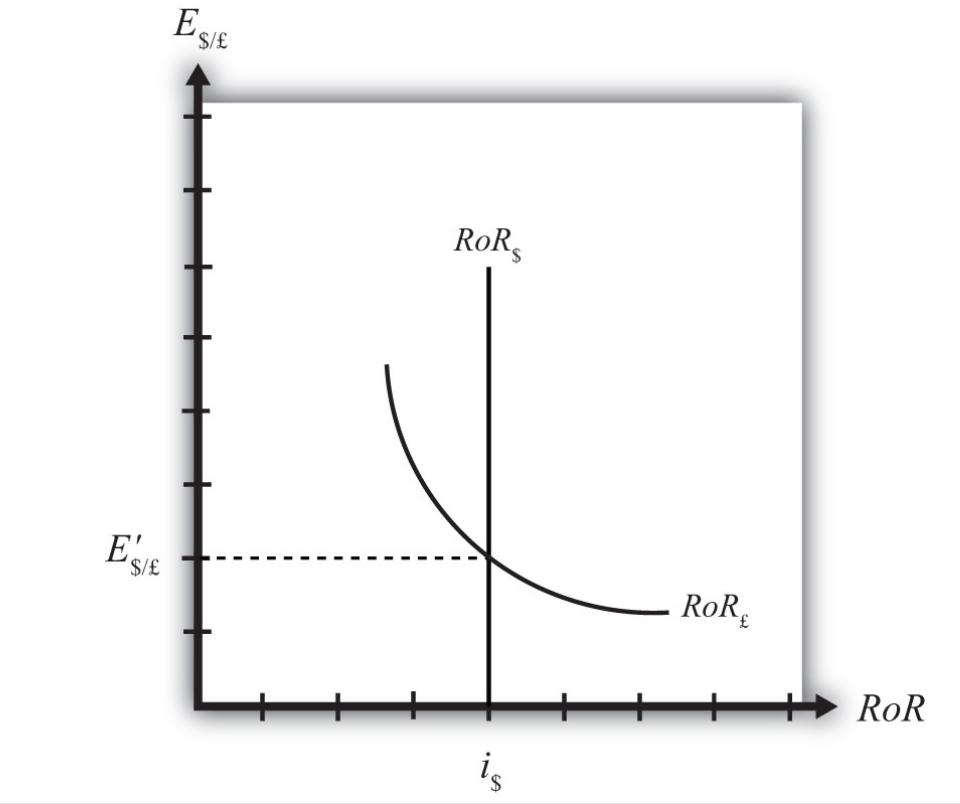
$$RoR_{\$} = i_{\$}.$$

The rate of return on the British asset, however, is a more complicated formula that depends on the British interest rate ($i_{\text{£}}$), the spot exchange rate ($E_{\$/\text{£}}$), and the expected exchange rate ($E_{\$/\text{£}}^e$). In its simplest form it is written as follows:

$$RoR_{\text{£}} = \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} (1 + i_{\text{£}}) - 1.$$

In [Figure 16.5 "Rate of Return Diagram"](#), we plot both RoR equations with respect to the exchange rate ($E_{\$/\text{£}}$). Since $RoR_{\$}$ is not a function (i.e., not dependent) on the exchange rate, it is drawn as a vertical line at the level of the U.S. interest rate ($i_{\$}$). This simply means that as the exchange rate rises or falls, the $RoR_{\$}$ always remains immutably fixed at the U.S. interest rate.

Figure 16.5 Rate of Return Diagram



The $\text{RoR}_{\text{£}}$, however, is a function of the exchange rate. Indeed, the relationship is negative since $E_{\$/\text{£}}$ is in the denominator of the equation. This means that as $E_{\$/\text{£}}$ rises, $\text{RoR}_{\text{£}}$ falls, and vice versa.

The intuition behind this negative relationship is obtained by looking at the alternative (equivalent) formula for $\text{RoR}_{\text{£}}$:

$$\text{RoR}_{\text{£}} = i_{\text{£}} + \frac{E_{\$/\text{£}}^e - E_{\$/\text{£}}}{E_{\$/\text{£}}} (1 + i_{\text{£}}).$$

Recall that the exchange rate ratio represents the expected percentage change in the value of the pound. Suppose, as an example, that this term were positive. That would mean the investor believes the pound will appreciate during the term of the investment. Furthermore, since it is an expected appreciation of the pound, it will add to the total rate of return on the British investment. Next, suppose the spot exchange rate ($E_{\$/\text{£}}$) rises today. Assuming ceteris paribus, as we always do in these exercises, the expected exchange rate remains fixed. That will mean the numerator

of the exchange rate expression will fall in value, as will the value of the entire expression. The interpretation of this change is that the investor's expected appreciation of the pound falls, which in turn lowers the overall rate of return. Hence, we get the negative relationship between the $\$/\text{£}$ exchange rate and $\text{RoR}_{\text{£}}$.

The intersection of the two RoR curves in the diagram identifies the unique exchange rate $E'_{\$/\text{£}}$ that equalizes rates of return between the two countries. This exchange rate is in equilibrium because any deviations away from interest rate parity (IRP) will motivate changes in investor behavior and force the exchange back to the level necessary to achieve IRP. The equilibrium adjustment story is next.

KEY TAKEAWAYS

- The rates of return are plotted with respect to the exchange rate. The domestic rate of return does not depend on the exchange rate and hence is drawn as a vertical line. The foreign rate of return is negatively related to the value of the foreign currency.
- The intersection of the rates of return identifies the exchange rate that satisfies the interest rate parity condition.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *positive, negative, or zero*, the relationship between the U.S. interest rate and the rate of return on U.S. assets.
 - b. Of *positive, negative, or zero*, the relationship between the exchange rate ($E_{\$/\text{£}}$) and the rate of return on U.S. assets.
 - c. Of *positive, negative, or zero*, the relationship between the exchange rate ($E_{\$/\text{£}}$) and the rate of return on British assets.
 - d. The name of the endogenous variable whose value is determined at the intersection of two rate of return curves.

16.4 Exchange Rate Equilibrium Stories with the RoR Diagram

LEARNING OBJECTIVE

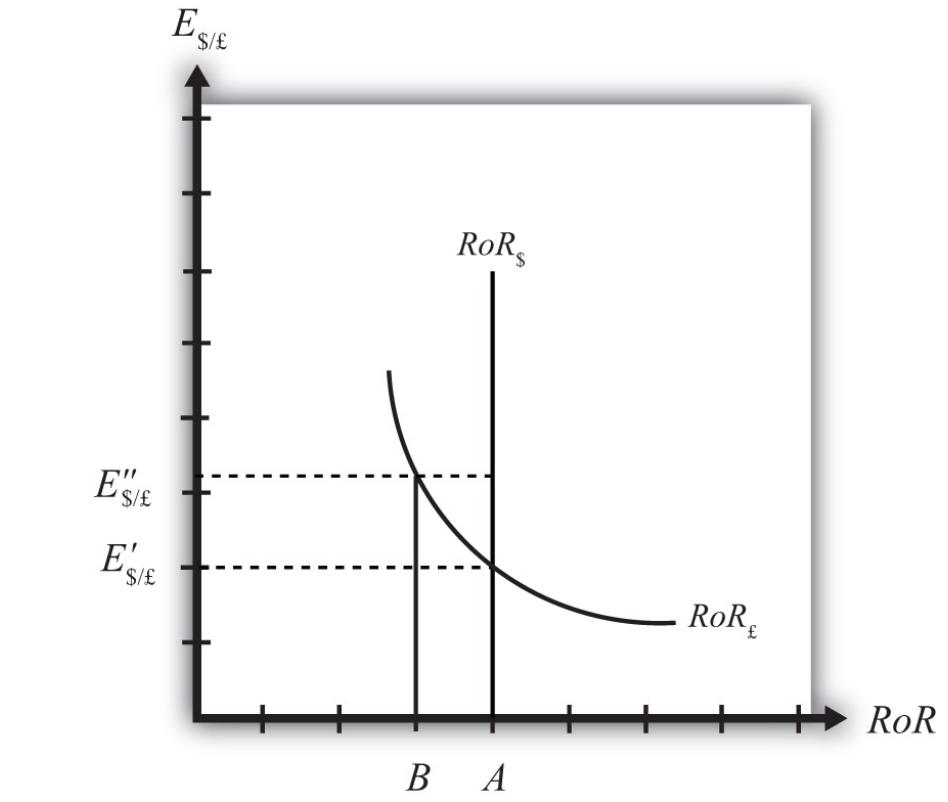
1. Learn how adjustment to equilibrium is described in the interest rate parity model.

Any equilibrium in economics has an associated behavioral story to explain the forces that will move the endogenous variable to the equilibrium value. In the foreign exchange (Forex) model, the endogenous variable is the exchange rate. This is the variable that is determined as a solution in the model and will change to achieve the equilibrium. Variables that do not change in the adjustment to the equilibrium are the exogenous variables. In this model, the exogenous variables are $E_{\$/\text{£}}^e$, $i_{\$}$, and $i_{\text{£}}$. Changes in the exogenous variables are necessary to cause an adjustment to a new equilibrium. However, in telling an equilibrium story, it is typical to simply assume that the endogenous variable is not at the equilibrium (for some unstated reason) and then explain how and why the variable will adjust to the equilibrium value.

Exchange Rate Too High

Suppose, for some unspecified reason, the exchange rate is currently at $E''_{\$/\text{£}}$ as shown in [Figure 16.6 "Adjustment When the Exchange Rate Is Too High"](#). The equilibrium exchange rate is at $E'_{\$/\text{£}}$ since at this rate, rates of return are equal and interest rate parity (IRP) is satisfied. Thus at $E''_{\$/\text{£}}$ the exchange rate is too high. Since the exchange rate, as written, is the value of the pound, we can also say that the pound value is too high relative to the dollar to satisfy IRP.

Figure 16.6 Adjustment When the Exchange Rate Is Too High



With the exchange rate at $E''_{\$/\u00a3}$, the rate of return on the dollar, $RoR_{\$}$, is given by the value A along the horizontal axis. This will be the value of the U.S. interest rate. The rate of return on the pound, $RoR_{\u00a3}$ is given by the value B , however. This means that $RoR_{\u00a3} < RoR_{\$}$ and IRP does not hold. Under this circumstance, higher returns on deposits in the United States will motivate investors to invest funds in the United States rather than Britain. This will raise the supply of pounds on the Forex as British investors seek the higher average return on U.S. assets. It will also lower the demand for British pounds ($\u00a3$) by U.S. investors who decide to invest at home rather than abroad. Both changes in the Forex market will lower the value of the pound and raise the U.S. dollar value, reflected as a reduction in $E_{\$/\u00a3}$.

In more straightforward terms, when the rate of return on dollar deposits is higher than on British deposits, investors will increase demand for the higher RoR currency and reduce demand for the other. The change in demand on the Forex raises the value of the currency whose RoR was initially higher (the U.S. dollar in this case) and lowers the other currency value (the British pound).

As the exchange rate falls from E''/\text{£}$ to E'/\text{£}$, $\text{RoR}_{\text{£}}$ begins to rise up, from B to A . This occurs because $\text{RoR}_{\text{£}}$ is negatively related to changes in the exchange rate. Once the exchange rate falls to E'/\text{£}$, $\text{RoR}_{\text{£}}$ will become equal to $\text{RoR}_{\$}$ at A and IRP will hold. At this point there are no further pressures in the Forex for the exchange rate to change, hence the Forex is in equilibrium at E'/\text{£}$.

Exchange Rate Too Low

If the exchange rate is lower than the equilibrium rate, then the adjustment will proceed in the opposite direction. At any exchange rate below E'/\text{£}$ in the diagram, $\text{RoR}_{\text{£}} > \text{RoR}_{\$}$. This condition will inspire investors to move their funds to Britain with the higher rate of return. The subsequent increase in the demand for pounds will raise the value of the pound on the Forex and $E_{\$/\text{£}}$ will rise (consequently, the dollar value will fall). The exchange rate will continue to rise and the rate of return on pounds will fall until $\text{RoR}_{\text{£}} = \text{RoR}_{\$}$ (IRP holds again) at E'/\text{£}$.

KEY TAKEAWAYS

- In the interest rate parity model, when the $$/\text{£}$ exchange rate is less than the equilibrium rate, the rate of return on British deposits exceeds the RoR on U.S. deposits. That inspires investors to demand more pounds on the Forex to take advantage of the higher RoR. Thus the $$/\text{£}$ exchange rate rises (i.e., the pound appreciates) until the equilibrium is reached when interest rate parity holds.
- In the interest rate parity model, when the $$/\text{£}$ exchange rate is greater than the equilibrium rate, the rate of return on U.S. deposits exceeds the RoR on British deposits. That inspires investors to demand more U.S. dollars on the Forex to take advantage of the higher RoR. Thus the $$/\text{£}$ exchange rate falls (i.e., the dollar appreciates) until the equilibrium is reached when interest rate parity holds.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the expected effect on the exchange rate ($E_{\$/\text{£}}$) if the rate of return on pound assets is greater than the rate of return on dollar assets.
 - b. Of *increase, decrease, or stay the same*, the expected effect on the exchange rate ($E_{\$/\text{£}}$) if the rate of return on U.S. assets is greater than the rate of return on British assets.
 - c. Of *increase, decrease, or stay the same*, the expected effect on the value of the dollar if the rate of return on pound assets is greater than the rate of return on dollar assets.
 - d. Of *increase, decrease, or stay the same*, the expected effect on the value of the dollar if the rate of return on U.S. assets is greater than the rate of return on British assets.
 - e. Of *increase, decrease, or stay the same*, the expected effect on the value of the dollar if the rate of return on U.S. assets is equal to the rate of return on British assets.

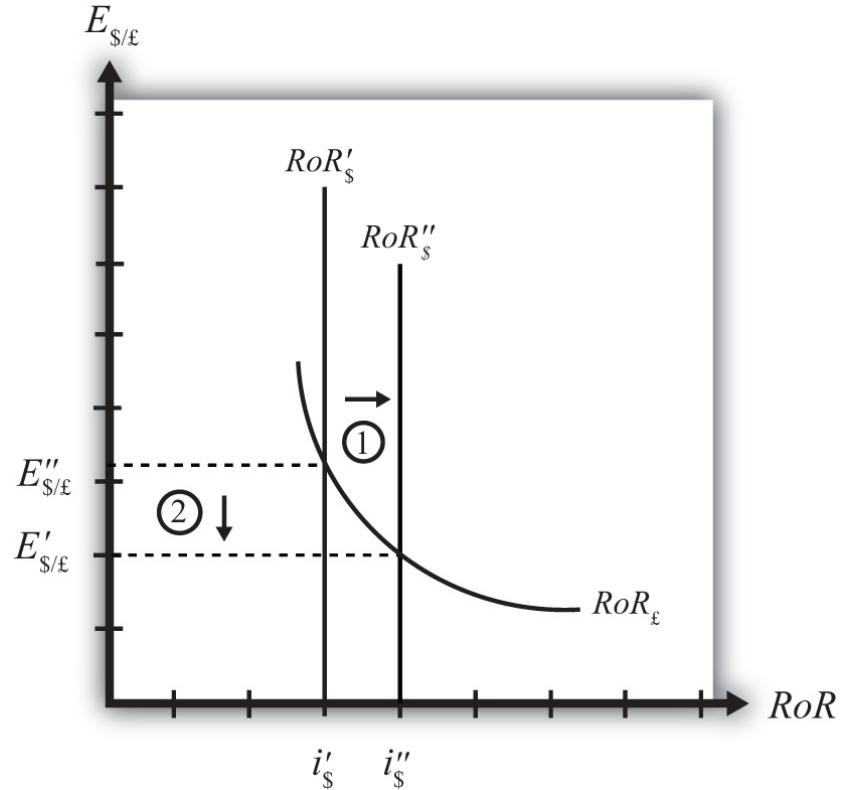
16.5 Exchange Rate Effects of Changes in U.S. Interest Rates Using the RoR Diagram

LEARNING OBJECTIVE

1. Learn the effects of changes in the foreign interest rate on the value of the domestic and foreign currency using the interest rate parity model.

Suppose that the foreign exchange market (Forex) is initially in equilibrium such that $RoR_E = RoR_{\$}$ (i.e., interest rate parity holds) at an initial equilibrium exchange rate given by $E'_{\$/\text{£}}$. The initial equilibrium is depicted in [Figure 16.7 "Effects of a U.S. Interest Rate Increase in a RoR Diagram"](#). Next, suppose U.S. interest rates rise, ceteris paribus. Ceteris paribus means we assume all other exogenous variables remain fixed at their original values. In this model, the British interest rate (i_E) and the expected exchange rate ($E_{\$/\text{£}}^e$) both remain fixed as U.S. interest rates rise.

Figure 16.7 Effects of a U.S. Interest Rate Increase in a RoR Diagram



The increase in U.S. interest rates will shift the U.S. RoR line to the right from $RoR'_{\$}$ to $RoR''_{\$}$ as indicated by step 1 in [Figure 16.7 "Effects of a U.S. Interest Rate Increase in a RoR Diagram"](#). Immediately after the increase and before the exchange rate changes, $RoR_{\$} > RoR_E$. The adjustment to the new equilibrium will follow the “exchange rate too high” equilibrium story earlier. Accordingly, higher U.S. interest rates will make U.S. dollar investments more attractive to investors, leading to an increase in demand for dollars on the Forex resulting in an appreciation of the dollar, a depreciation of the pound, and a decrease in $E_{\$/\text{£}}$. The exchange rate will fall to the new equilibrium rate $E''_{\$/\text{£}}$ as indicated by step 2 in the figure.

In summary, an increase in the U.S. interest rate will raise the rate of return on dollars above the rate of return on pounds, lead investors to shift investments to U.S. assets, and result in a decrease in the $\$/\text{£}$ exchange rate (i.e., an appreciation of the U.S. dollar and a depreciation of the British pound).

In contrast, a decrease in U.S. interest rates will lower the rate of return on dollars below the rate of return on pounds, lead investors to shift investments to British assets, and result in an increase in the $\$/\text{£}$ exchange rate (i.e., a depreciation of the U.S. dollar and an appreciation of the British pound).

KEY TAKEAWAYS

- An increase in U.S. interest rates will result in a decrease in the $\$/\text{£}$ exchange rate (i.e., an appreciation of the U.S. dollar and a depreciation of the British pound).
- A decrease in U.S. interest rates will result in an increase in the $\$/\text{£}$ exchange rate (i.e., a depreciation of the U.S. dollar and an appreciation of the British pound).

EXERCISE

1. Consider the economic change listed along the top row of the following table. In the empty boxes, indicate the effect of each change, sequentially, on the variables listed in the first column. For example, a decrease in U.S. interest rates will cause a decrease in the rate of return (RoR) on U.S. assets. Therefore a “-” is placed in the first cell under the “A Decrease in U.S. Interest Rates” column of the table. Next in sequence, answer how the RoR on euro assets will be affected. Use the interest rate parity model to determine the answers. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	A Decrease in U.S. Interest Rates
RoR on U.S. Assets	-
RoR on Euro Assets	
Demand for U.S. Dollars on the Forex	
Demand for Euros on the Forex	
U.S. Dollar Value	
Euro Value	
E/\epsilon$	

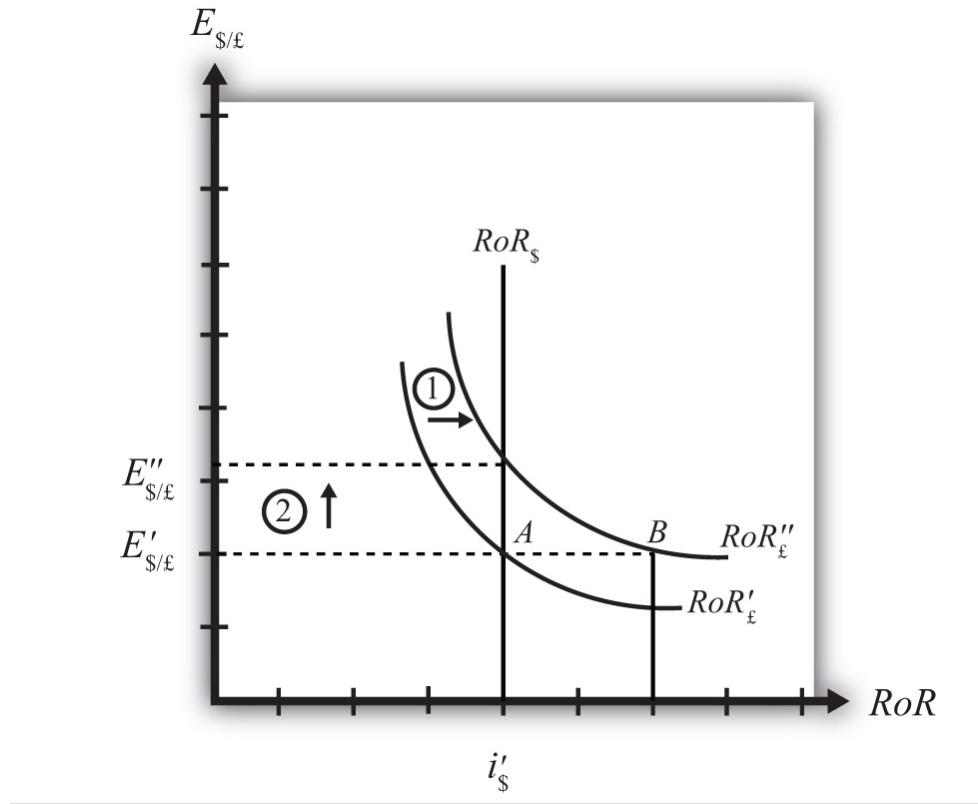
16.6 Exchange Rate Effects of Changes in Foreign Interest Rates Using the RoR Diagram

LEARNING OBJECTIVE

1. Learn the effects of changes in the foreign interest rate on the value of the domestic and foreign currency using the interest rate parity model.

Suppose that the foreign exchange market (Forex) is initially in equilibrium such that $RoR_E = RoR_\$/\text{£}$ (i.e., interest rate parity holds) at an initial equilibrium exchange rate given by $E' \$/\text{£}$. The initial equilibrium is depicted in [Figure 16.8 "Effects of a British Interest Rate Increase in a RoR Diagram"](#). Next, suppose British interest rates rise, *ceteris paribus*. *Ceteris paribus* means we assume all other exogenous variables remain fixed at their original values. In this model, the U.S. interest rate ($i_\$$) and the expected exchange rate ($E_\$/\text{£}^e$) both remain fixed as British interest rates rise.

Figure 16.8 Effects of a British Interest Rate Increase in a RoR Diagram



The increase in British interest rates ($i_\£$) will shift the British RoR line to the right from $RoR'_\£$ to $RoR''_\£$ as indicated by step 1 in the figure.

The reason for the shift can be seen by looking at the simple rate of return formula:

$$RoR_\£ = \frac{E_{\$/\£}^e}{E_{\$/\£}} (1 + i_\£) - 1.$$

Suppose one is at the original equilibrium with exchange rate $E'_{\$/\£}$. Looking at the formula, an increase in $i_\£$ clearly raises the value of $RoR_\£$ for any fixed values of $E_{\$/\£}^e$. This could be represented as a shift to the right on the diagram, as from A to B. Once at B with a new interest rate, one could perform the exercise used to plot out the downward sloping RoR curve (see [Chapter 16 "Interest Rate Parity", Section 16.3 "Forex Equilibrium with the Rate of Return Diagram"](#)). The result would be a curve, like the original, but shifted entirely to the right.

Immediately after the increase and before the exchange rate changes, $RoR_E > RoR_\$/\text{£}$. The adjustment to the new equilibrium will follow the “exchange rate too low” equilibrium story presented in [Chapter 16 "Interest Rate Parity", Section 16.4 "Exchange Rate Equilibrium Stories with the RoR Diagram"](#). Accordingly, higher British interest rates will make British pound investments more attractive to investors, leading to an increase in demand for pounds on the Forex, and resulting in an appreciation of the pound, a depreciation of the dollar, and an increase in $E\$/\text{£}$. The exchange rate will rise to the new equilibrium rate $E''\$/\text{£}$ as indicated by step 2.

In summary, an increase in British interest rates will raise the rate of return on pounds above the rate of return on dollars, lead investors to shift investments to British assets, and result in an increase in the $\$/\text{£}$ exchange rate (i.e., an appreciation of the British pound and a depreciation of the U.S. dollar).

In contrast, a decrease in British interest rates will lower the rate of return on British pounds below the rate of return on dollars, lead investors to shift investments to U.S. assets, and result in a decrease in the $\$/\text{£}$ exchange rate (i.e., a depreciation of the British pound and an appreciation of the U.S. dollar).

KEY TAKEAWAYS

- An increase in British interest rates will result in an increase in the $\$/\text{£}$ exchange rate (i.e., an appreciation of the British pound and a depreciation of the U.S. dollar).
- A decrease in British interest rates will result in a decrease in the $\$/\text{£}$ exchange rate (i.e., a depreciation of the British pound and an appreciation of the U.S. dollar).

EXERCISE

1. Consider the economic change listed along the top row of the following table. In the empty boxes, indicate the effect of each change, sequentially, on the variables listed in the first column. For example, a decrease in U.S. interest rates will cause a decrease in the rate of return (RoR) on U.S. assets. Therefore a “-” is placed in the first box of the table. Next in sequence, answer how the RoR on euro assets will be affected. Use the interest rate parity model to determine the answers. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	A Decrease in Euro Interest Rates
RoR on U.S. Assets	-
RoR on Euro Assets	
Demand for U.S. Dollars on the Forex	
Demand for Euros on the Forex	
U.S. Dollar Value	
Euro Value	
E/\epsilon$	

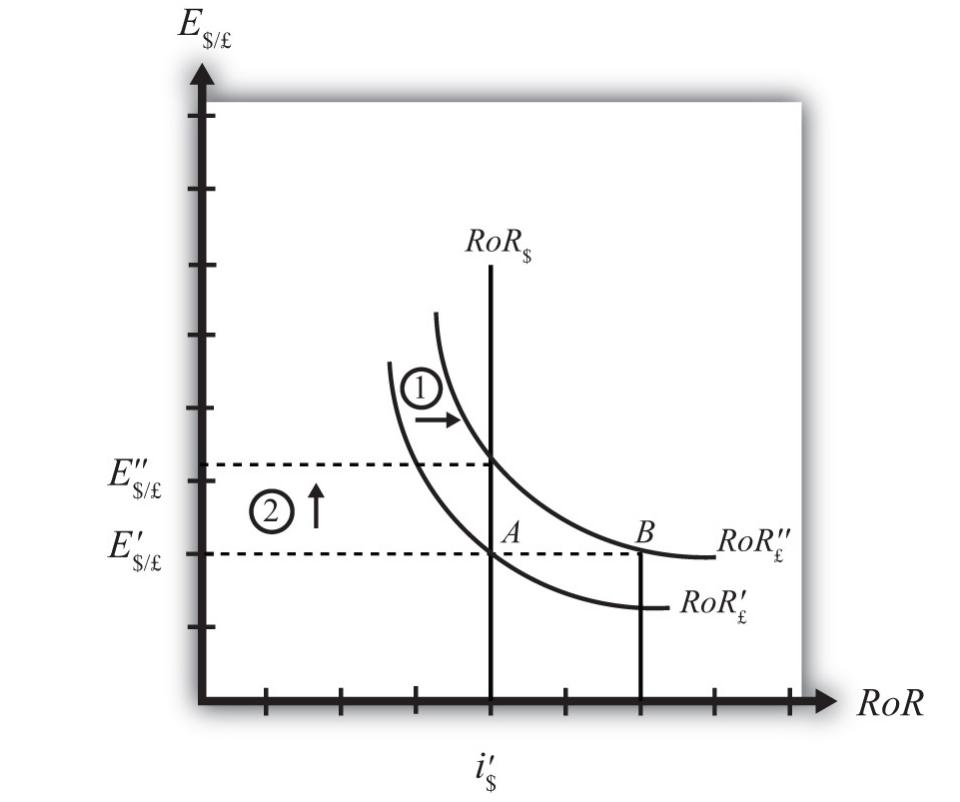
16.7 Exchange Rate Effects of Changes in the Expected Exchange Rate Using the RoR Diagram

LEARNING OBJECTIVE

1. Learn the effects of changes in the expected future currency value on the spot value of the domestic and foreign currency using the interest rate parity model.

Suppose that the foreign exchange market (Forex) is initially in equilibrium such that $RoR_E = RoR_\$/\text{£}$ (i.e., interest rate parity holds) at an initial equilibrium exchange rate given by $E' \$/\text{£}$. The initial equilibrium is depicted in [Figure 16.9 "Effects of an Expected Exchange Rate Change in a RoR Diagram"](#). Next, suppose investors' beliefs shift so that $E \$/\text{£}^e$ rises, ceteris paribus. Ceteris paribus means we assume all other exogenous variables remain fixed at their original values. In this model, the U.S. interest rate ($i_\$$) and the British interest rate (i_E) both remain fixed as the expected exchange rate rises.

Figure 16.9 Effects of an Expected Exchange Rate Change in a RoR Diagram



An expected exchange rate increase means that if investors had expected the pound to appreciate, they now expect it to appreciate even more. Likewise, if investors had expected the dollar to depreciate, they now expect it to depreciate more. Alternatively, if they had expected the pound to depreciate, they now expect it to depreciate less. Likewise, if they had expected the dollar to appreciate, they now expect it to appreciate less.

This change might occur because new information is released. For example, the British Central Bank might release information that suggests an increased chance that the pound will rise in value in the future.

The increase in the expected exchange rate ($E_{\$/\£}^e$) will shift the British RoR line to the right from $RoR'_{\£}$ to $RoR''_{\£}$ as indicated by step 1 in the figure.

The reason for the shift can be seen by looking at the simple rate of return formula:

$$RoR_{\text{£}} = \frac{E_{\$/\text{£}}^e}{E_{\$/\text{£}}} (1 + i_{\text{£}}) - 1.$$

Suppose one is at the original equilibrium with exchange rate $E'_{\$/\text{£}}$. Looking at the formula, an increase in $E_{\$/\text{£}}^e$ clearly raises the value of $RoR_{\text{£}}$ for any fixed values of $i_{\text{£}}$. This could be represented as a shift to the right on the diagram from A to B. Once at B with a new expected exchange rate, one could perform the exercise used to plot out the downward sloping RoR curve. The result would be a curve, like the original, but shifted entirely to the right.

Immediately after the increase and before the exchange rate changes, $RoR_{\text{£}} > RoR_{\$}$. The adjustment to the new equilibrium will follow the “exchange rate too low” equilibrium story presented in [Chapter 16 "Interest Rate Parity", Section 16.4 "Exchange Rate Equilibrium Stories with the RoR Diagram"](#). Accordingly, higher expected British rates of return will make British pound investments more attractive to investors, leading to an increase in demand for pounds on the Forex and resulting in an appreciation of the pound, a depreciation of the dollar, and an increase in $E_{\$/\text{£}}$. The exchange rate will rise to the new equilibrium rate $E''_{\$/\text{£}}$ as indicated by step 2.

In summary, an increase in the expected future $\$/\text{£}$ exchange rate will raise the rate of return on pounds above the rate of return on dollars, lead investors to shift investments to British assets, and result in an increase in the $\$/\text{£}$ exchange rate (i.e., an appreciation of the British pound and a depreciation of the U.S. dollar).

In contrast, a decrease in the expected future $\$/\text{£}$ exchange rate will lower the rate of return on British pounds below the rate of return on dollars, lead investors to shift investments to U.S. assets, and result in a decrease in the $\$/\text{£}$ exchange rate (i.e., a depreciation of the British pound and an appreciation of the U.S. dollar).

KEY TAKEAWAYS

- An increase in the expected future pound value (with respect to the U.S. dollar) will result in an increase in the spot $\$/\text{£}$ exchange rate (i.e., an appreciation of the British pound and a depreciation of the U.S. dollar).
- A decrease in the expected future pound value (with respect to the U.S. dollar) will result in a decrease in the spot $\$/\text{£}$ exchange rate (i.e., a depreciation of the British pound and an appreciation of the U.S. dollar).

EXERCISE

1. Consider the economic change listed along the top row of the following table. In the empty boxes, indicate the effect of the change, sequentially, on the variables listed in the first column. For example, a decrease in U.S. interest rates will cause a decrease in the rate of return (RoR) on U.S. assets. Therefore a “-” is placed in the first box of the table. Next in sequence, answer how the RoR on euro assets will be affected. Use the interest rate parity model to determine the answers. You do not need to show your work. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	A Reduction in Next Year's Expected Dollar Value
RoR on U.S. Assets	-
RoR on Euro Assets	
Demand for U.S. Dollars on the Forex	
Demand for Euros on the Forex	
U.S. Dollar Value	
Euro Value	
E/\epsilon$	

Chapter 17

Purchasing Power Parity

Purchasing power parity is both a theory about exchange rate determination and a tool to make more accurate comparisons of data between countries. It is probably more important in its latter role since as a theory it performs pretty poorly. Its poor performance arises largely because its simple form depends on several assumptions that are not likely to hold in the real world and because the amount of foreign exchange activity due to importer and exporter demands is much less than the amount of activity due to investor demands. Nonetheless, the theory remains important to provide the background for its use as a tool for cross-country comparisons of income and wages, which is used by international organizations like the World Bank in presenting much of their international data.

17.1 Overview of Purchasing Power Parity (PPP)

LEARNING OBJECTIVES

1. Identify the conditions under which the law of one price holds.
2. Identify the conditions under which purchasing power parity holds.

Purchasing power parity (PPP)¹ is a theory of exchange rate determination and a way to compare the average costs of goods and services between countries. The theory assumes that the actions of importers and exporters (motivated by cross-country price differences) induce changes in the spot exchange rate. In another vein, PPP suggests that transactions on a country's current account affect the value of the exchange rate on the foreign exchange (Forex) market. This is in contrast with the interest rate parity theory, which assumes that the actions of investors (whose transactions are recorded on the capital account) induce changes in the exchange rate.

PPP theory is based on an extension and variation of the “law of one price” as applied to the aggregate economy. To explain the theory it is best to first review the idea behind the law of one price.

The Law of One Price (LoOP)

The law of one price says that identical goods should sell for the same price in two separate markets when there are no transportation costs and no differential taxes applied in the two markets. Consider the following information about movie video tapes sold in the U.S. and Mexican markets.

Price of videos in U.S. market ($P_{\v)	\$20
Price of videos in Mexican market (P_p^v)	p150
Spot exchange rate ($E_{p/\$}$)	10 p/\$

1. A theory of exchange rate determination based on traders' motivations that result in a PPP exchange rate when there are no transportation costs and no differential taxes applied.

The dollar price of videos sold in Mexico can be calculated by dividing the video price in pesos by the spot exchange rate as shown:

$$\frac{P_p^v}{E_{p/\$}} \left[\frac{\text{peso/video}}{\text{peso}/\$} = \frac{\text{peso}}{\text{video}} \times \frac{\$}{\text{peso}} = \frac{\$}{\text{video}} \right] = \frac{150}{10} = \$15/\text{video}.$$

To see why the peso price is divided by the exchange rate rather than multiplied, notice the conversion of units shown in the brackets. If the law of one price held, then the dollar price in Mexico should match the price in the United States. Since the dollar price of the video is less than the dollar price in the United States, the law of one price *does not hold* in this circumstance.

The next question to ask is what might happen as a result of the discrepancy in prices. Well, as long as there are no costs incurred to transport the goods, there is a profit-making opportunity through trade. For example, U.S. travelers in Mexico who recognize that identical video titles are selling there for 25 percent less might buy videos in Mexico and bring them back to the United States to sell. This is an example of “goods arbitrage.” An arbitrage opportunity arises whenever one can buy something at a low price in one location, resell it at a higher price, and thus make a profit.

Using basic supply and demand theory, the increase in demand for videos in Mexico would push up the price of videos. The increase in supply of videos on the U.S. market would force the price down in the United States. In the end, the price of videos in Mexico may rise to, say, p₁₈₀ while the price of videos in the United States may fall to \$18. At these new prices *the law of one price holds* since

$$\frac{P_p^v}{E_{p/\$}} = \frac{180}{10} = \$18 = P_{\$}^v.$$

The idea in the law of one price is that identical goods selling in an integrated market in which there are no transportation costs, no differential taxes or subsidies, and no tariffs or other trade barriers should sell at identical prices. If different prices prevailed, then there would be profit-making opportunities by buying the good in the low price market and reselling it in the high price market. If entrepreneurs took advantage of this arbitrage opportunity, then the prices would converge to equality.

Of course, for many reasons the law of one price does not hold even between markets within a country. The price of beer, gasoline, and stereos will likely be different in New York City and in Los Angeles. The price of these items will also be different in other countries when converted at current exchange rates. The simple reason for the discrepancies is that there are costs to transport goods between locations, there are different taxes applied in different states and different countries, nontradable input prices may vary, and people do not have perfect information about the prices of goods in all markets at all times. Thus to refer to this as an economic “law” does seem to exaggerate its validity.

From LoOP to PPP

The purchasing power parity theory is really just the law of one price applied in the aggregate but with a slight twist added. If it makes sense from the law of one price that identical goods should sell for identical prices in different markets, then the law ought to hold for all identical goods sold in both markets.

First, let's define the variable $CB_{\$}$ to represent the cost of a basket of goods in the United States denominated in dollars. For simplicity we could imagine using the same basket of goods used in the construction of the U.S. consumer price index ($CPI_{\$}$). The **consumer price index (CPI)**² uses a market basket of goods that are purchased by an average household during a specified period. The basket is determined by surveying the quantity of different items purchased by many different households. One can then determine, on average, how many units of bread, milk, cheese, rent, electricity, and so on are purchased by the typical household. You might imagine it's as if all products are purchased in a grocery store with items being placed in a basket before the purchase is made. $CB_{\$}$ then represents the dollar cost of purchasing all the items in the market basket. We will similarly define CB_p to be the cost of a market basket of goods in Mexico denominated in pesos.

Now if the law of one price holds for each individual item in the market basket, then it should hold for the market baskets as well. In other words,

$$\frac{P_p^v}{E_{p/\$}} = P_{\$}^v \Rightarrow \frac{CB_p}{E_{p/\$}} = CB_{\$}.$$

Rewriting the right-hand side equation allows us to put the relationship in the form commonly used to describe absolute purchasing power parity, which is

$$E_{p/\$}^{PPP} = \frac{CB_p}{CB_{\$}}.$$

- 2. An index that measures the average level of prices of goods and services in an economy relative to a base year.
- 3. The exchange rate that equalizes the cost of a market basket of goods between two countries.

If this condition holds between two countries, then we would say PPP is satisfied. The condition says that the **PPP exchange rate**³ (pesos per dollar) will equal the ratio of the costs of the two market baskets of goods denominated in local currency units. Note that the reciprocal relationship $E_{\$/p}^{PPP} = \frac{CB_{\$}}{CB_p}$ is also valid.

Because the cost of a market basket of goods is used in the construction of the country's consumer price index, PPP is often written as a relationship between the

exchange rate and the country's price indices. However, it is not possible merely to substitute the price index directly for the cost of the market basket used above. To see why, we will review the construction of the CPI in [Chapter 17 "Purchasing Power Parity"](#), [Section 17.2 "The Consumer Price Index \(CPI\) and PPP"](#).

KEY TAKEAWAYS

- The law of one price says that identical goods should sell for identical prices in two different markets when converted at the current exchange rate and when there are no transportation costs and no differential taxes applied.
- The purchasing power parity theory is an aggregated version of the law of one price.
- The purchasing power parity condition says that identical market baskets should sell for identical prices in two different markets when converted at the current exchange rate and when there are no transportation costs and no differential taxes applied.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The exchange rate value if toothpaste costs \$2.50 in the United States and 30 pesos in Mexico and the law of one price holds.
 - b. The exchange rate value if a market basket costs \$450 in the United States and 5,400 pesos in Mexico and purchasing power parity holds.
 - c. The term used to describe a collection of goods and services consumed by a typical consumer.
 - d. The term used to distinguish PPP based on price levels rather than inflation rates.
 - e. The term used to describe the economic principle that identical goods should sell at identical prices in different markets.
2. Use the information in the table below to answer the following question. Show your work.

	The Economist Price per Issue	Exchange Rate (December 2, 1999)
United States	\$3.95	-
Canada	C\$ 4.95	1.47 C\$/ \$
Japan	¥920	102 ¥/\$

- a. Calculate the implied purchasing power parity exchange rates between Canada and the United States and between Japan and the United States based on the price of the *Economist* magazine.

17.2 The Consumer Price Index (CPI) and PPP

LEARNING OBJECTIVE

1. Learn the relationship between the consumer price index and the PPP exchange rate.

The consumer price index (CPI) is an index that measures the average level of prices of goods and services in an economy relative to a base year. To track only what happens to prices, the quantities of goods purchased is assumed to remain fixed from year to year. This is accomplished by determining—with survey methods—the average quantities of all goods and services purchased by a typical household during some period. The quantities of all of these goods together are referred to as the average market basket. For example, the survey might find that the average household in one month purchases 10 gallons of gas, 15 cans of beer, 3.2 gallons of milk, 2.6 pounds of butter, and so on. The basket of goods would also contain items like health and auto insurance, housing services, utility services, and many other items. We can describe the market basket easily as a collection or set of quantities ($Q_1, Q_2, Q_3, \dots, Q_n$). Here Q_1 may be the quantity of gasoline, Q_2 the quantity of beer, and so on. The set has n different quantity entries, implying that there are n different items in the market basket.

The cost of the market basket is found by surveying the average prices for each of the n products in the market in question. This survey would yield a collection or set of prices ($P_1, P_2, P_3, \dots, P_n$). The cost of the market basket is then found by summing the product of the price and quantity for each item. That is, $CB = P_1Q_1 + P_2Q_2 + P_3Q_3 + \dots + P_nQ_n$, or

$$CB = \sum_{i=1}^n P_iQ_i.$$

The first year in which the index is constructed is called the base year. Suppose 1982 is the base year for the United States. Let CB_{YY} represent the cost of the market basket evaluated at the prices that prevail in the year (YY) (e.g., CB_{09} is the cost of a market basket evaluated in 2009 prices). The CPI is derived according to the following formula:

$$CPI_{YY} = \frac{CB_{YY}}{CB_{82}} \times 100,$$

where CPI_{YY} is the CPI in the year (YY). The term is multiplied by 100 by convention, probably because it reduces the need to use digits after a decimal point. Notice that the CPI in the base year is equal to 100—that is, $CPI_{82} = 100$ —because $CB_{82}/CB_{82} = 1$. This is true for all indices—they are by convention set to 100 in the base year.

The CPI in a different year (either earlier or later) represents the ratio of the cost of the market basket in that year relative to the cost of the same basket in the base year. If in 1982 the cost of the market basket rises, then the CPI will rise above 100. If the cost of the market basket falls, then the CPI would fall below 100.

If the CPI rises, it does not mean that the prices of all the goods in the market basket have risen. Some prices may rise more or less. Some prices may even fall. The CPI measures the average price change of goods and services in the basket.

The inflation rate for an economy is the percentage change in the CPI during a year. Thus if CPI_{08} on January 1, 2008, and CPI_{09} on January 1, 2009, are the price indices, then the inflation rate during 2008 is given by

$$\pi_{08} = \frac{CPI_{09} - CPI_{08}}{CPI_{08}} \times 100.$$

PPP Using the CPI

The purchasing power parity relationship can be written using the CPI with some small adjustments. First, consider the following ratio of 2009 consumer price indices between Mexico and the United States:

$$\frac{CPI_p^{09}}{CPI_{\$}^{09}} = \frac{CB_p^{09}/CB_p^{08}}{CB_{\$}^{09}/CB_{\$}^{08}} = \frac{CB_p^{09}}{CB_{\$}^{09}} \frac{CB_{\$}^{08}}{CB_p^{08}} = \frac{CB_p^{09}/CB_{\$}^{09}}{CB_p^{08}/CB_{\$}^{08}}.$$

Given that the base year is 2008, the ratio is written in terms of the market basket costs on the right-hand side and then rewritten into another form. The far right-hand side expression now reflects the purchasing power parity exchange rates in 2009 divided by the PPP exchange rate in 2008, the base year. In other words,

$$\frac{CPI_p^{09}}{CPI_{\$}^{08}} = \frac{CB_p^{09}/CB_{\$}^{09}}{CB_p^{08}/CB_{\$}^{09}} = \frac{^{09}E_{p/\$}^{\text{PPP}}}{^{08}E_{p/\$}^{\text{PPP}}}.$$

So, in general, if you want to use the consumer price indices for two countries to derive the PPP exchange rate for 2009, you must apply the following formula, derived by rewriting the above as

$$^{09}E_{p/\$}^{\text{PPP}} = {}^{\text{base}}E_{p/\$}^{\text{PPP}} \times \frac{CPI_p^{09}}{CPI_{\$}^{09}},$$

where ${}^{\text{base}}E_{p/\$}^{\text{PPP}}$ represents the PPP exchange rate that prevails in the base year between the two countries. Note that in order for this formula to work correctly, the CPIs in both countries must share the same base year. If they did not, a more complex formula would need to be derived.

KEY TAKEAWAYS

- A country's consumer price index in year (YY) is derived as the ratio of the market basket cost in year (YY) and the market basket cost in the base year.
- The PPP exchange rate between two countries can be written as the ratio of their consumer price indices in that year multiplied by an adjustment factor given by the PPP exchange rate in the base year of the countries' CPIs.

EXERCISE

1. Suppose a consumer purchases the following products each week: ten gallons of gas, fifteen cans of beer, three gallons of milk, and two pounds of butter. Suppose in the initial week the prices of the products are \$3 per gallon of gas, \$2 per can of beer, \$4 per gallon of milk, and \$4 per pound of butter. Suppose one year later the prices of the same products are \$2 per gallon of gas, \$3 per can of beer, \$5 per gallon of milk, and \$5 per pound of butter.
 - a. Calculate the cost of a weekly market basket in the initial base period.
 - b. Calculate the cost of a market basket one year later.
 - c. Construct the price index value for both years.
 - d. What is the inflation rate between the two years?

17.3 PPP as a Theory of Exchange Rate Determination

LEARNING OBJECTIVE

1. Learn how adjustment to equilibrium occurs in the PPP model.

The purchasing power parity (PPP) relationship becomes a theory of exchange rate determination by introducing assumptions about the behavior of importers and exporters in response to changes in the relative costs of national market baskets. Recall the story of the law of one price, when the price of a good differed between two countries' markets and there was an incentive for profit-seeking individuals to buy the good in the low price market and resell it in the high price market. Similarly, if a market basket containing many different goods and services costs more in one market than another, we should likewise expect profit-seeking individuals to buy the relatively cheaper goods in the low-cost market and resell them in the higher-priced market. If the law of one price leads to the equalization of the prices of a good between two markets, then it seems reasonable to conclude that PPP, describing the equality of market baskets across countries, should also hold.

However, adjustment within the PPP theory occurs with a twist compared to adjustment in the law of one price story. In the law of one price story, goods arbitrage in a particular product was expected to affect the prices of the goods in the two markets. The twist that's included in the PPP theory is that arbitrage, occurring across a range of goods and services in the market basket, will affect the exchange rate rather than the market prices.

PPP Equilibrium Story

To see why the PPP relationship represents an equilibrium, we need to tell an equilibrium story. An equilibrium story in an economic model is an explanation of how the behavior of individuals will cause the equilibrium condition to be satisfied. The equilibrium condition is the PPP equation written as

$$E_{p/\$}^{\text{PPP}} = \frac{CB_p}{CB_\$} .$$

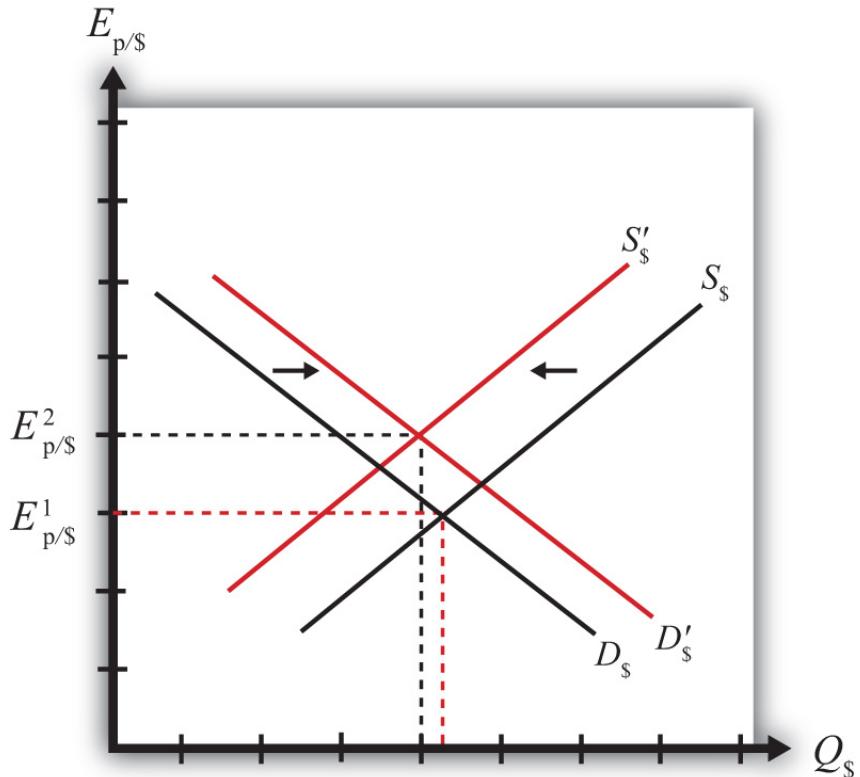
The endogenous variable in the PPP theory is the exchange rate. Thus we need to explain why the exchange rate will change if it is not in equilibrium. In general there are always two versions of an equilibrium story, one in which the endogenous variable ($E_p/\$$ here) is too high and one in which it is too low.

PPP equilibrium story 1. Let's consider the case in which the exchange rate is too low to be in equilibrium. This means that

$$E_{p/\$} < \frac{CB_p}{CB_{\$}} \Rightarrow CB_{\$}E_{p/\$} < CB_p,$$

where $E_p/\$$ is the exchange rate that prevails on the spot market. Since it is less than the ratio of the market basket costs in Mexico and the United States, it is also less than the PPP exchange rate. The right side of the expression is rewritten to show that the cost of a market basket in the United States evaluated in pesos (i.e., $CB_{\$}E_{p/\$}$) is less than the cost of the market basket in Mexico also evaluated in pesos. Thus it is cheaper to buy the basket in the United States, or in other words, it is more profitable to sell items in the market basket in Mexico.

The PPP theory now suggests that the cheaper basket in the United States will lead to an increase in demand for goods in the U.S. market basket by Mexico. As a consequence, it will increase the demand for U.S. dollars on the foreign exchange (Forex) market. Dollars are needed because purchases of U.S. goods require U.S. dollars. Alternatively, U.S. exporters will realize that goods sold in the United States can be sold at a higher price in Mexico. If these goods are sold in pesos, the U.S. exporters will want to convert the proceeds back to dollars. Thus there is an increase in U.S. dollar demand (by Mexican importers) and an increase in peso supply (by U.S. exporters) on the Forex. This effect is represented by a rightward shift in the U.S. dollar demand curve in [Figure 17.1 "Forex Adjustment When"](#). At the same time, U.S. consumers will reduce their demand for the pricier Mexican goods. This will reduce the supply of dollars (in exchange for pesos) on the Forex, which is represented by a leftward shift in the U.S. dollar supply curve in the Forex market.

Figure 17.1 Forex Adjustment When $E_{p/\$}$ Is Low


Both the shift in demand and supply will cause an increase in the value of the dollar and thus the exchange rate ($E_{p/\$}$) will rise. As long as the U.S. market basket remains cheaper, excess demand for the dollar will persist and the exchange rate will continue to rise. The pressure for change ceases once the exchange rate rises enough to equalize the cost of market baskets between the two countries and PPP holds.

PPP equilibrium story 2. Now let's consider the other equilibrium story (i.e., the case in which the exchange rate is too high to be in equilibrium). This implies that

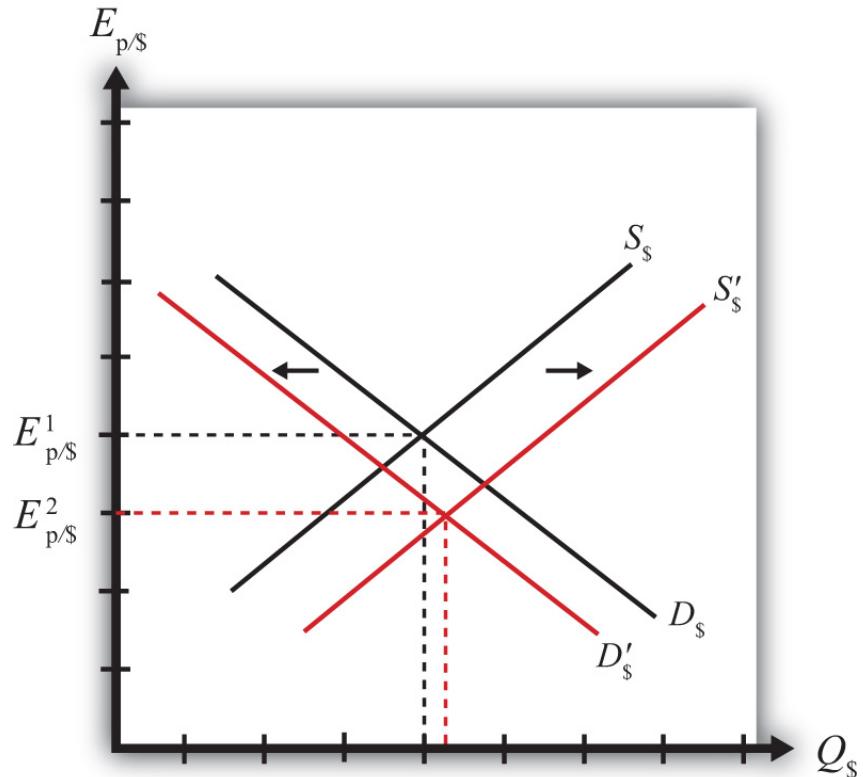
$$E_{p/\$} > \frac{CB_p}{CB_{\$}} \Rightarrow CB_{\$}E_{p/\$} > CB_p.$$

The left-side expression says that the spot exchange rate is greater than the ratio of the costs of market baskets between Mexico and the United States. In other words, the exchange rate is above the PPP exchange rate. The right-side expression says that the cost of a U.S. market basket, converted to pesos at the current exchange rate, is greater than the cost of a Mexican market basket in pesos. Thus, on average,

U.S. goods are relatively more expensive while Mexican goods are relatively cheaper.

The price discrepancies should lead consumers in the United States or importing firms to purchase less expensive goods in Mexico. To do so, they will raise the supply of dollars in the Forex in exchange for pesos. Thus the supply curve of dollars will shift to the right as shown in [Figure 17.2 "Forex Adjustment When"](#). At the same time, Mexican consumers would refrain from purchasing the more expensive U.S. goods. This would lead to a reduction in demand for dollars in exchange for pesos on the Forex. Hence, the demand curve for dollars shifts to the left. Due to the demand decrease and the supply increase, the exchange rate ($E_{p/\$}$) falls. This means that the dollar depreciates and the peso appreciates.

Figure 17.2 Forex Adjustment When $E_{p/\$}$ Is High



Extra demand for pesos will continue as long as goods and services remain cheaper in Mexico. However, as the peso appreciates (the dollar depreciates), the cost of Mexican goods rises relative to U.S. goods. The process ceases once the PPP exchange rate is reached and market baskets cost the same in both markets.

Adjustment to Price Level Changes under PPP

In the PPP theory, exchange rate changes are induced by changes in relative price levels between two countries. This is true because the quantities of the goods are always presumed to remain fixed in the market baskets. Therefore, the only way that the cost of the basket can change is if the goods' prices change. Since price level changes represent inflation rates, this means that differential inflation rates will induce exchange rate changes according to the theory.

If we imagine that a country begins with PPP, then the inequality given in equilibrium story 1, $CB_{\$}E_{p/\$} < CB_p$, can arise if the price level rises in Mexico (peso inflation), if the price level falls in the United States (dollar deflation), or if Mexican inflation is more rapid than U.S. inflation. According to the theory, the behavior of importers and exporters would now induce a dollar appreciation and a peso depreciation. In summary, an increase in Mexican prices relative to the change in U.S. prices (i.e., more rapid inflation in Mexico than in the United States) will cause the dollar to appreciate and the peso to depreciate according to the purchasing power parity theory.

Similarly, if a country begins with PPP, then the inequality given in equilibrium story 2, $CB_{\$}E_{p/\$} > CB_p$, can arise if the price level rises in the United States (dollar inflation), the price level falls in Mexico (peso deflation), or if U.S. inflation is more rapid than Mexican inflation. In this case, the inequality would affect the behavior of importers and exporters and induce a dollar depreciation and peso appreciation. In summary, more rapid inflation in the United States would cause the dollar to depreciate while the peso would appreciate.

KEY TAKEAWAYS

- An increase in Mexican prices relative to the change in U.S. prices (i.e., more rapid inflation in Mexico than in the United States) will cause the dollar to appreciate and the peso to depreciate according to the purchasing power parity theory.
- More rapid inflation in the United States would cause the dollar to depreciate while the peso would appreciate.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or no change*, the effect on the demand for euros in the foreign exchange market if a market basket costs more in the United States than it does in Germany.
 - b. Of *increase, decrease, or no change*, the effect on the supply of dollars in the foreign exchange market if a market basket costs more in the United States than it does in Germany.
 - c. Of *increase, decrease, or no change*, the effect on the U.S. dollar value according to the PPP theory if a market basket costs \$300 in the United States and €200 in Germany and the exchange rate is $E\$/\epsilon = 1.30$.
 - d. Of *increase, decrease, or no change*, the effect on the euro value according to the PPP theory if a market basket costs €200 in Germany and ¥22,000 in Japan and the exchange rate is $E¥/\epsilon = 115$.
 - e. Of *increase, decrease, or no change*, the effect on the euro value according to the PPP theory if a market basket costs €200 in Germany and ¥22,000 in Japan and the exchange rate is $E¥/\epsilon = 100$.

17.4 Problems and Extensions of PPP

LEARNING OBJECTIVES

1. Identify the reasons why the PPP condition is rarely satisfied between two countries.
2. Learn the dynamic version of PPP.

Problems with the PPP Theory

The main problem with the purchasing power parity (PPP) theory is that the PPP condition is rarely satisfied within a country. There are quite a few reasons that can explain this and so, given the logic of the theory, which makes sense, economists have been reluctant to discard the theory on the basis of lack of supporting evidence. Below we consider some of the reasons PPP may not hold.

Transportation costs and trade restrictions. Since the PPP theory is derived from the law of one price, the same assumptions are needed for both theories. The law of one price assumes that there are no transportation costs and no differential taxes applied between the two markets. These mean that there can be no tariffs on imports or other types of restrictions on trade. Since transport costs and trade restrictions do exist in the real world, this would tend to drive prices for similar goods apart. Transport costs should make a good cheaper in the exporting market and more expensive in the importing market. Similarly, an import tariff would drive a wedge between the prices of an identical good in two trading countries' markets, raising it in the import market relative to the export market price. Thus the greater transportation costs and trade restrictions are between countries, the less likely for the costs of market baskets to be equalized.

Costs of nontradable inputs. Many items that are homogeneous nevertheless sell for different prices because they require a nontradable input in the production process. As an example, consider why the price of a McDonald's Big Mac hamburger sold in downtown New York City is higher than the price of the same product in the New York suburbs. Because the rent for restaurant space is much higher in the city center, the restaurant will pass along its higher costs in the form of higher prices. Substitute products in the city center (other fast food restaurants) will face the same high rental costs and thus will charge higher prices as well. Because it would be impractical (i.e., costly) to produce the burgers at a cheaper suburban location

and then transport them for sale in the city, competition would not drive the prices together in the two locations.

Perfect information. The law of one price assumes that individuals have good, even perfect, information about the prices of goods in other markets. Only with this knowledge will profit seekers begin to export goods to the high price market and import goods from the low-priced market. Consider a case in which there is imperfect information. Perhaps some price deviations are known to traders but other deviations are not known, or maybe only a small group of traders know about a price discrepancy and that group is unable to achieve the scale of trade needed to equalize the prices for that product. (Perhaps they face capital constraints and can't borrow enough money to finance the scale of trade needed to equalize prices.) In either case, traders without information about price differences will not respond to the profit opportunities and thus prices will not be equalized. Thus the law of one price may not hold for some products, which would imply that PPP would not hold either.

Other market participants. Notice that in the PPP equilibrium stories, it is the behavior of profit-seeking importers and exporters that forces the exchange rate to adjust to the PPP level. These activities would be recorded on the current account of a country's balance of payments. Thus it is reasonable to say that the PPP theory is based on current account transactions. This contrasts with the interest rate parity theory in which the behavior of investors seeking the highest rates of return on investments motivates adjustments in the exchange rate. Since investors are trading assets, these transactions would appear on a country's capital account of its balance of payments. Thus the interest rate parity theory is based on capital account transactions.

It is estimated that there are approximately \$1–2 trillion dollars worth of currency exchanged every day on international foreign exchange (Forex) markets. That's one-eighth of U.S. GDP, which is the value of production in the United States in an entire year. In addition, the \$1–2 trillion estimate is made by counting only one side of each currency trade. Thus that's an enormous amount of trade. If one considers the total amount of world trade each year and then divides by 365, one can get the average amount of goods and services traded daily. This number is less than \$100 billion dollars. This means that the amount of daily currency transactions is more than ten times the amount of daily trade. This fact would seem to suggest that the primary effect on the daily exchange rate must be caused by the actions of investors rather than importers and exporters. Thus the participation of other traders in the Forex market, who are motivated by other concerns, may lead the exchange rate to a value that is not consistent with PPP.

Relative PPP

There is an alternative version of the PPP theory called the “**relative PPP theory**⁴. In essence this is a dynamic version of the absolute PPP theory. Since absolute PPP suggests that the exchange rate may respond to inflation, we can imagine that the exchange rate would change in a systematic way given that a continual change in the price level (inflation) is occurring.

In the relative PPP theory, exchange rate changes over time are assumed to be dependent on inflation rate differentials between countries according to the following formula:

$$\frac{E_{p/\$}^2 - E_{p/\$}^1}{E_{p/\$}^1} = \pi_p - \pi_\$.$$

Here the percentage change in the dollar value between the first period and the second period is given on the left side. The right side gives the differences in the inflation rates between Mexico and the United States that were evaluated over the same time period. The implication of relative PPP is that if the Mexican inflation rate exceeds the U.S. inflation rate, then the dollar will appreciate by that differential over the same period. The logic of this theory is the same as in absolute PPP. Importers and exporters respond to variations in the relative costs of market baskets so as to maintain the law of one price, at least on average. If prices continue to rise faster in Mexico than in the United States, for example, price differences between the two countries would grow and the only way to keep up with PPP is for the dollar to appreciate continually versus the peso.

KEY TAKEAWAYS

- Purchasing power parity (PPP) will not be satisfied between countries when there are transportation costs, trade barriers (e.g., tariffs), differences in prices of nontradable inputs (e.g., rental space), imperfect information about current market conditions, and when other Forex market participants, such as investors, trade currencies for other reasons.
- Relative PPP is a dynamic version of the theory that relates currency appreciation or depreciation to differences in country inflation rates.

4. A dynamic version of the PPP theory that relates currency appreciation or depreciation to differences in inflation rates between countries.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The name for the PPP theory based on relative inflation rates between countries.
 - b. A type of trade cost whose presence is likely to cause deviations in the law of one price and PPP.
 - c. The term used to describe a kind of production input, of which office rental is one type.
 - d. Traders need to have information about this in other markets in order to take advantage of arbitrage opportunities.

17.5 PPP in the Long Run

LEARNING OBJECTIVE

1. Interpret the PPP theory as a projection of long-term tendencies in exchange rate values.

In general, the purchasing power parity (PPP) theory works miserably when applied to real-world data. In other words, it is rare for the PPP relationship to hold true between any two countries at any particular point in time. In most scientific disciplines, the failure of a theory to be supported by the data means the theory is refuted and should be thrown out or tossed away. However, economists have been reluctant to do that with the PPP theory. In part this is because the logic of the theory seems particularly sound. In part it's because there are so many "frictions" in the real world, such as tariffs, nontariff barriers, transportation costs, measurement problems, and so on that it would actually be surprising for the theory to work when applied directly to the data. (It is much like expecting an object to follow Newton's laws of motion while sitting on the ground.)

In addition, economists have conceived of an alternative way to interpret or apply the PPP theory to overcome the empirical testing problem. The trick is to think of PPP as a "long-run" theory of exchange rate determination rather than a short-run theory. Under such an interpretation, it is no longer necessary for PPP to hold at any point in time. Instead, the PPP exchange rate is thought to represent a target toward which the spot exchange rate is slowly drawn.

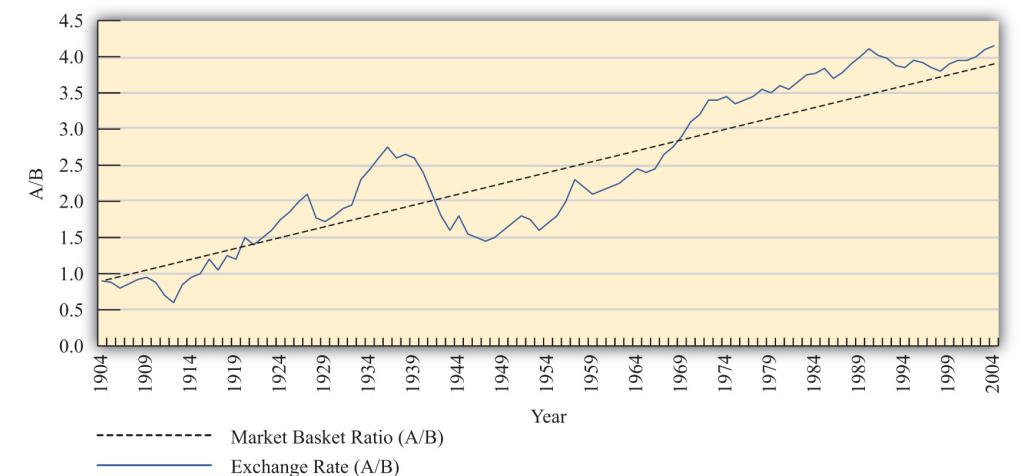
This long-run interpretation requires an assumption that importers and exporters cannot respond quickly to deviations in the cost of market baskets between countries. Instead of immediate responses to price differences between countries by engaging in arbitrage—buying at the low price and selling high—traders respond slowly to these price signals. Some reasons for the delay include imperfect information (traders are not aware of the price differences), long-term contracts (traders must wait till current contractual arrangements expire), and/or marketing costs (entry to new markets requires research and setup costs). In addition, we recognize that the exchange rate is not solely determined by trader behavior. Investors, who respond to different incentives, might cause persistent deviations from the PPP exchange rate even if traders continue to respond to the price differences.

When there is a delayed response, PPP no longer needs to hold at a particular point in time. However, the theory does imagine that traders eventually will adjust to the price differences (buying low and selling high), causing an eventual adjustment of the spot exchange rate toward the PPP rate. However, as adjustment occurs, it is quite possible that the PPP exchange rate also continues to change. In this case, the spot exchange rate is adjusting toward a moving target.

How long will this adjustment take? In other words, how long is the long run? The term itself is generally used by economists to represent some “unspecified” long period of time; it might be several months, years, or even decades. Also, since the target, the PPP exchange rate, is constantly changing, it is quite possible that it is never reached. The adjustment process may never allow the exchange rate to catch up to the target even though it is constantly chasing it.

Perhaps the best way to see what the long-run PPP theory suggests is to consider [Figure 17.3 "Hypothetical Long-Term Trend"](#). The figure presents constructed data (i.e., made up) between two countries, A and B. The dotted black line shows the ratio of the costs of market baskets between the two countries over a long period, a century between 1904 and 2004. It displays a steady increase, indicating that prices have risen faster in country A relative to country B. The solid blue line shows a plot of the exchange rate between the two countries during the same period. If PPP were to hold at every point in time, then the exchange rate plot would lie directly on top of the market basket ratio plot. The fact that it does not means PPP did not hold all the time. In fact, PPP held only at times when the exchange rate plot crosses the market basket ratio plot; on the diagram this happened only twice during the century—not a very good record.

Figure 17.3 Hypothetical Long-Term Trend

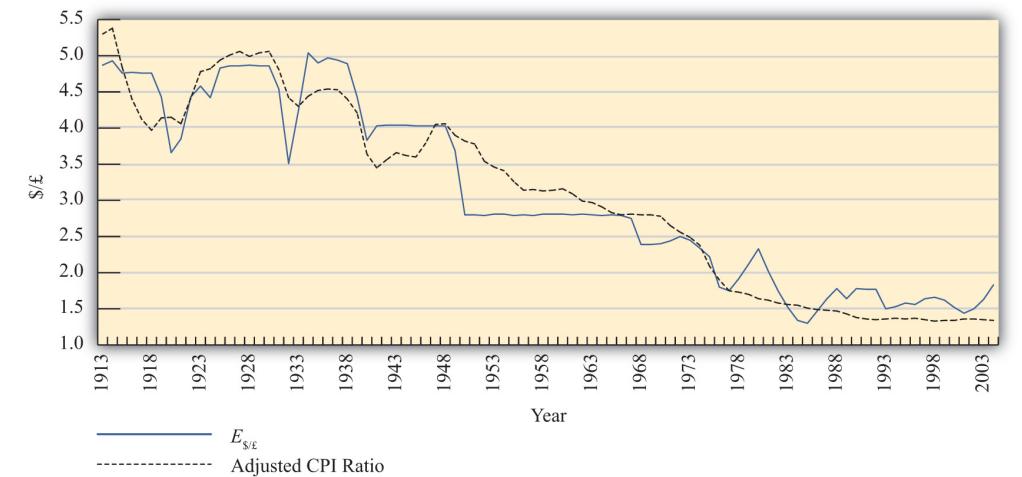


Nonetheless, despite performing poorly with respect to moment-by-moment PPP, the figure displays an obvious regularity. The trend of the exchange rate between the countries is almost precisely the trend in the market basket ratio; both move upward at about the same “average” rate. Sometimes the exchange rate is below the market basket ratio, even for a long period of time, but at other times, the exchange rate rises up above the market basket ratio.

The idea here is that lengthy exchange rate deviations from the market basket ratio (i.e., the PPP exchange rate) mean long periods of time in which the cost of goods is cheaper in one country than in another. Eventually, traders will respond to these price discrepancies and begin to import more from the less expensive country. This will lead to the increase in demand for that country’s currency and cause the exchange rate to move back toward the market basket ratio. However, in the long-run version of the theory, this will take time, sometimes a considerable amount of time, even years or more.

To see how this relationship works in one real-world example, consider [Figure 17.4 "U.S./UK Long-Term Trends"](#). It plots the exchange rate ($E_{\$/\text{£}}$) between the U.S. dollar and the British pound between 1913 and 2004 together with an adjusted ratio of the countries’ consumer price indices (CPIs) during the same period. A technical point: The ratio of CPIs is adjusted because the ratio of CPIs must be multiplied by the PPP exchange rate that prevailed in the base year for the two countries. However, the CPI series used has 1967 as the base year in the United Kingdom and 1974 as the base year in the United States. This would mean the CPI ratio should be multiplied by the ratio of the cost of a market basket in the United States in 1974 divided by the market basket cost in the United Kingdom in 1967. Unsurprisingly, I don’t have that information. Thus I’ll assume a number (1.75) that is somewhat greater than the actual exchange rate that prevailed at the time. The higher number may account for the fact that prices rose considerably between 1967 and 1974. In any case, it remains a guess. The adjusted ratio represents an estimate of the ratio of the costs of market baskets between the two countries.

Figure 17.4 U.S./UK Long-Term Trends



In the diagram, the dotted black line represents the estimated ratio of market basket costs and the solid blue line is the exchange rate ($E_{\$/\text{£}}$). Note how closely the exchange rate tracks the trend in the market basket ratio. This remains true even though the exchange rate remained fixed during some lengthy periods of time, as in the 1950s and 1960s. While this depiction is just two countries over a long period, it is suggestive that the long-run version of PPP may have some validity.

More sophisticated empirical tests of the long-run version of PPP have shown mixed results, as some studies support the hypothesis while others tend to reject it. Regardless, there is much more support for this version of the theory than for the much more simplistic short-run version.

KEY TAKEAWAYS

- Under the long-run purchasing power parity (PPP) theory, the PPP exchange rate is thought to represent a target toward which the spot exchange rate is slowly drawn over time. The empirical evidence for this theory is mixed.
- Long-run data showing the trend in consumer price index (CPI) ratios between the United States and the United Kingdom relative to the $\$/\text{£}$ exchange rate suggest some validity to the theory.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used by economists to denote an unspecified point in time in the distant future.
 - b. The term used by economists to denote an unspecified point in time in the near future.
 - c. The term used to describe the general path along which a variable is changing.
 - d. Under this version of the PPP theory, the PPP exchange rate represents a target toward which the spot exchange rate is slowly drawn over time.

17.6 Overvaluation and Undervaluation

LEARNING OBJECTIVE

1. Recognize how the terms *overvalued* and *undervalued* exchange rates are defined, applied, and interpreted.

It is quite common to hear people claim that a country's exchange rate is overvalued or undervalued. The first question one should ask when someone claims the exchange rate is overvalued is "overvalued with respect to what?" There are two common reference exchange rates often considered. The person may mean the exchange rate is overvalued with respect to purchasing power parity (PPP), or he may mean the exchange rate is overvalued relative to the rate presumed needed to balance the current account (CA).

The mere use of these terms suggests immediately that there is some "proper" value for the exchange rate. However, one should refrain from accepting this implication. As was previously discussed, PPP is unlikely to hold, even over very long periods, for a variety of very good reasons. Also, there is no reason to think that current account balance represents some equilibrium or goal for an economy: countries can run trade deficits or surpluses for an extended period and suffer no ill effects. Thus overvaluation or undervaluation of an exchange rate, for either reason (PPP or current account balance) should be thought of simply as something that happens. Of more interest is what it means when it happens.

Over- and Undervaluation with Respect to PPP

First let's consider over- and undervaluation with respect to PPP. The PPP exchange rate is defined as the rate that equalizes the cost of a market basket of goods between two countries. The PPP exchange rate between the Mexican peso and the U.S. dollar would be written as

$$E_{p/\$}^{\text{PPP}} = \frac{CB_p}{CB_\$},$$

which represents the PPP value of the U.S. dollar in terms of pesos.

If the U.S. dollar is overvalued with respect to the Mexican peso, then the spot exchange rate exceeds the PPP exchange rate:

$$E_{p/\$} > E_{p/\$}^{\text{PPP}}.$$

This will also mean the exchange rate exceeds the ratio of market basket costs:

$$E_{p/\$} > \frac{CB_p}{CB_{\$}};$$

therefore, the following will hold:

$$E_{p/\$} CB_{\$} > CB_p.$$

The left side (LS) of this expression represents the cost of a U.S. market basket converted to pesos at the current spot exchange rate. The right side (RS) is the cost of the basket in Mexico also evaluated in pesos. Since $LS > RS$, goods and services cost more on average in the United States than in Mexico at the current exchange rate. Thus for the U.S. dollar to be overvalued with respect to the peso means that goods and services are relatively more expensive in the United States than in Mexico. Of course, it also implies that goods and services are relatively cheaper in Mexico.

A simple guide to judge whether a currency is overvalued is to consider it from the perspective of a tourist. When the U.S. dollar is overvalued, a U.S. tourist traveling to Mexico will find that many products seem cheaper than in the United States, after converting at the spot exchange rate. Thus an overvalued currency will buy more in other countries.

An undervalued currency works in the opposite direction. When the U.S. dollar is undervalued, the cost of a basket of goods in the United States is lower than the cost in Mexico when evaluated at the current exchange rate. To a U.S. tourist, Mexican goods and services would seem more expensive on average. Thus an undervalued currency will buy less in other countries.

Finally, if the U.S. dollar is overvalued with respect to the Mexican peso, it follows that the peso is undervalued with respect to the dollar. In this case, since the U.S. tourists would find Mexican goods comparatively cheap, Mexican tourists would find U.S. goods to be comparatively expensive. If the U.S. dollar were undervalued, then the peso would be overvalued.

Is overvaluation or undervaluation good or bad? That depends on what a person is trying to achieve. For example, if the U.S. dollar is overvalued with respect to the peso, then a U.S. tourist traveling to Mexico will be very happy. In fact, the more overvalued the dollar is, the better. However, for an exporter of U.S. goods to Mexico, its price in peso terms will be higher the more overvalued is the dollar. Thus an overvalued dollar will likely reduce sales and profits for these U.S. firms.

Over- and Undervaluation with Respect to Current Account Balance

The second way over- and undervaluation is sometimes applied is in comparison to an exchange rate presumed necessary to induce trade balance, or balance on the current account. If one imagines that a trade deficit, for example, arises primarily because a country imports too much or exports too little (rather than being driven by financial decisions tending to cause a financial account surplus), then one may also look for ways to either reduce imports or raise exports. A change in the exchange rate offers one viable method to affect trade flows.

Suppose the United States has a trade deficit (which it indeed has had for more than thirty years prior to 2010). If the U.S. dollar value were to fall—a dollar depreciation—then foreign goods would all become relatively more expensive to U.S. residents, tending to reduce U.S. imports. At the same time, a dollar depreciation would also cause U.S. goods to become relatively cheaper to foreign residents tending to raise U.S. exports.

Sometimes economists make numerical estimations as to how much the dollar value would have to fall to bring trade into balance. These estimations are enormously difficult to make for several reasons and should be interpreted and used with great caution, if at all. The primary reason is that many different factors on both the trade side and the financial side influence a country's trade imbalance besides just the exchange rate. The exchange rate that balances trade would depend on the values taken by all the other factors that also influence the trade balance. Different values for all the other variables would mean a different exchange rate needed to balance trade. Thus there isn't *one* exchange rate value that will balance trade. Instead, there is a different exchange rate value that will balance trade in each and every alternative circumstance. Indeed, even the current exchange rate—whatever that is—can balance trade if other factors change appropriately.

Despite these cautions, many observers will still contend that a country's currency needs to depreciate by some percentage to eliminate a trade deficit, or needs to appreciate to eliminate a trade surplus. When it is believed a depreciation of the currency is needed to balance trade, they will say the currency is overvalued. When

it is believed an appreciation of the currency is needed to balance trade, they will say the currency is undervalued. However, in a floating exchange rate system, it is hard to argue that the exchange rate is at the “wrong” value since—with competition in the market—it will always be at the rate that equalizes supply and demand. In other words, the “proper” value for the exchange rate can be said to be *not* the one that will satisfy PPP or not the one that will generate trade balance but rather whatever rate currently prevails. Under this notion, a currency can never be over- or undervalued in a floating exchange rate system. Instead, the spot exchange rate is always at the “proper” value.

In a fixed exchange rate system, a government can sometimes intervene to maintain an exchange rate that is very different from what would arise if allowed to float. In these cases, large trade surpluses can arise because the government maintains an artificially low value for its currency. Calls for a revaluation (appreciation) of the currency, to promote a reduction in a trade surplus, are somewhat more appropriate in these cases since the market does not determine the exchange rate. Similarly, large deficits could be reduced with a devaluation (depreciation) of the currency.

KEY TAKEAWAYS

- A currency can be overvalued or undervalued with respect to two reference values: (1) the value that would satisfy purchasing power parity (PPP) or (2) the value that would generate current account balance.
- Use of the terms *overvaluation* and *undervaluation* suggests that there is a “proper” value for the exchange rate. However, there are often valid reasons why exchange rates will not conform to PPP or why trade imbalances will persist.
- In a floating exchange rate system, the “proper” exchange rate can be said to be the rate that equalizes supply and demand for currencies in exchange. Under this notion, there can never be an over- or undervalued exchange rate.

EXERCISES

1. Use the information in the table below to answer the question, “Is the U.S. dollar overvalued or undervalued with respect to the Canadian dollar and the Japanese yen in terms of purchases of the *Economist*?” State why it is overvalued or undervalued. Show your work.

	The <i>Economist</i> Price per Issue	Exchange Rate (December 2, 1999)
United States	\$3.95	-
Canada	C\$4.95	1.47 C\$/ \$
Japan	¥920	102 ¥/\$

2. Use the information in the table below to answer the following questions:

	Big Mac Price	Exchange Rate (June 4, 1998)
United States (dollar)	\$2.53	-
South Korea (won)	W 2,600	1,475 W/\$
Israel (shekel)	sh 12.50	3.70 sh/\$
Poland (zloty)	zl 5.30	3.46 zl/\$

- Calculate whether the won, shekel, and zloty are overvalued or undervalued with respect to the U.S. dollar in terms of Big Mac purchases. Explain what it means to be overvalued or undervalued.
- What would the exchange rates have to be in order to equalize Big Mac prices between South Korea and the United States, Israel and the United States, and Poland and the United States?
- If in the long run the exchange rate moves to satisfy Big Mac purchasing power parity (PPP), will the won, shekel, and

zloty appreciate or depreciate in terms of dollars? Explain the logic.

3. Use the information about the hourly wage for a high school principal and exchange rates to answer the following questions:

	Wage	Actual Exchange Rate	PPP Exchange Rate
United States	\$25/hour	-	-
Mexico	P220/hour	10.9 p/\$	7.5 p/\$
Japan	¥3,000/hour	110 ¥/\$	132 ¥/\$

- a. Calculate the hourly wage rate in dollars in Mexico and Japan using the actual exchange rates.
- b. Calculate the hourly wage rate in dollars in Mexico and Japan using the PPP exchange rates.
- c. Based on the information above, in which country is it best to be a high school principal? Which country is second best? Which is third best?
- d. In terms of PPP, is the U.S. dollar overvalued or undervalued with respect to the peso and with respect to the yen?
- e. According to the PPP theory, given the conditions above, would the dollar be expected to appreciate or depreciate with respect to the peso and with respect to the yen?

17.7 PPP and Cross-Country Comparisons

LEARNING OBJECTIVE

1. Learn why using PPP exchange rates to convert income data to a common currency is a better method for making cross-country comparisons.

Probably the most important application of purchasing power parity (PPP) exchange rates is in making cross-country comparisons of income, wages, or gross domestic product (GDP). Suppose that we would like to compare per capita GDP between two countries—say, the United States and China. In 2004, GDP in the United States was approximately \$12 trillion; in China GDP was about ¥16 trillion. With a population in the United States of 290 million people, per capita U.S. GDP works out to \$41,400 per person. China's population was approximately 1.3 billion people in 2004, so its GDP per capita works out to 11,500 yuan (¥) per person. However, we can't compare these two per capita figures since they are in different units—dollars and yuan. Thus we need to convert units, either turn dollars into yuan or yuan into dollars.

The simplest approach to make this conversion is to use the spot exchange rate that prevailed in 2004, which was 8.28 yuan per dollar. Converting yuan to dollars yields a per capita GDP for China of \$1,390. Note that at \$41,400 per person, U.S. per capita GDP was almost thirty times higher than China's.

However, there is a problem using this method. One thing that is quickly recognized by Americans when they travel in and around China is that many goods and services seem considerably cheaper than they are in the United States. From a Chinese traveler's perspective, many U.S. goods would seem considerably more expensive. The implication is that although U.S. GDP per person is thirty times higher, that income may not purchase thirty times more goods and services in the U.S. because the prices of U.S. goods and services are so much higher when converted at the current exchange rate. Since presumably we are comparing per capita GDPs to compare how “well-off” people are in one country relative to another, these per capita figures will not accurately reflect these differences.

A solution is found in the purchasing power parity theory (PPP). When prices for similar goods differ as described in the previous paragraph, we would say the U.S. dollar is overvalued with respect to the yuan and with respect to PPP. At the same

time, we would say the yuan is undervalued vis-à-vis the dollar. One way to reach comparable (or equalized) values of goods and services between the countries is to apply the PPP exchange rate in the conversion. The PPP exchange rate is that exchange rate that would equalize the value of comparable market baskets of goods and services between two countries.

For example, the estimated PPP exchange rate between the U.S. dollar and yuan in 2004 was 1.85 ¥/\$. If this exchange rate had prevailed between the countries, the prices of U.S. goods would seem, on average, to be approximately equal to the prices that prevailed in China. Now, if we use this exchange rate to make the conversion to dollars of GDP per capita in China, then we will get a number that reflects the purchasing power of Chinese income in terms of the prices that prevail in the United States—that is, in terms of prices that are equalized between the countries.

Thus if we take China's GDP per capita of ¥11,500 and convert to dollars with the PPP exchange rate, we get \$6,250 per person. The units derived in this expression would typically be called "international dollars." What this means is that ¥11,500 will buy a bundle of goods and services in China that would cost \$6,250 if purchased in the United States at U.S. prices. In other words, ¥11,500 is equal to \$6,250 when the prices of goods and services are equalized between countries.

The PPP method of conversion is a much more accurate way of making cross-country comparisons of values between countries. In this example, although China's per capita GDP was still considerably lower than in the United States (\$6,250 vs. \$41,400), it is nonetheless four and a half times higher than using the spot exchange rate (\$6,250 vs. \$1,390). The higher value takes account of the differences in prices between the countries and thus better reflects the differences in purchasing power of per capita GDP.

The PPP conversion method has become the standard method used by the World Bank and others in making cross-country comparisons of GDP, GDP per capita, and average incomes and wages. For most comparisons concerning the size of economies or standards of living, using PPP is a more accurate method and can fundamentally change our perception of how countries compare. To see how, consider [Table 17.1 "GDP Rankings \(in Billions of Dollars\), 2008"](#), constructed from World Bank data. It shows a ranking of the top ten countries in total GDP converting to dollars using both the current exchange rate method and the PPP method.

Table 17.1 GDP Rankings (in Billions of Dollars), 2008

Rank	Country	Using Current Exchange Rate (\$)	Country	Using PPP Exchange Rate (\$)
1	United States	14,204	United States	14,204
2	Japan	4,909	China	7,903
3	China	4,326	Japan	4,355
4	Germany	3,653	India	3,338
5	France	2,853	Germany	2,925
6	United Kingdom	2,646	Russia	2,288
7	Italy	2,293	United Kingdom	2,176
8	Brazil	1,613	France	2,112
9	Russia	1,608	Brazil	1,977
10	Spain	1,604	Italy	1,841
11	Canada	1,400	Mexico	1,542
12	India	1,217	Spain	1,456

The United States remains at the top of the list using both methods. However, several countries rise up in the rankings. China rises from the third largest economy using current exchange rates to the second largest using PPP. This means that in terms of the physical goods and services produced by the economies, China really does produce more than Japan. PPP conversion gives a better representation of the relative sizes of these countries.

Similarly, India rises from twelfth rank to fourth. Russia also moves up into sixth place from ninth. At the same time, Japan, Germany, the United Kingdom, France, Italy, Brazil, and Spain all move down in the rankings. Canada moves out of the top twelve, being replaced by Mexico, which rises up to eleventh.

For those countries whose GDP rises in value when converting by PPP (i.e., China, India, and Russia), their currencies are undervalued with respect to the U.S. dollar. So using the current exchange rate method underestimates the true size of their economies. For the other countries, their currencies are overvalued to the dollar, so

converting their GDPs at current exchange rates gives an overestimate of the true size of their economies.

KEY TAKEAWAYS

- Using purchasing power parity (PPP) exchange rates to convert income data to a common currency is a better way to make international comparisons because it compensates for the differential costs of living.
- “International dollars” is the term used for the units for data converted to U.S. dollars using the PPP exchange rate.
- International rankings can vary significantly between data converted using actual versus PPP exchange rates.

EXERCISES

1. In February 2004, the Mexican peso–U.S. dollar exchange rate was 11 p/\$. The price of a hotel room in Mexico City was 1,000 pesos. The price of a hotel room in New York City was \$200.
 - a. Calculate the price of the Mexican hotel room in terms of U.S. dollars.
 - b. Calculate the price of the U.S. hotel room in terms of Mexican pesos.
 - c. Now suppose the exchange rate rises to 12 p/\$. What does the exchange rate change indicate has happened to the value of the U.S. dollar relative to the value of the Mexican peso?
 - d. Does the currency change benefit the U.S. tourist traveling to Mexico City or the Mexican tourist traveling to New York City? Explain why.
2. In 2008, Brazil’s per capita income in nominal terms was \$8,295 while its per capita income in purchasing power parity (PPP) terms was \$10,466. Based on this information, if you were an American traveling in Brazil, would Brazilian products seem expensive or inexpensive relative to U.S. products?
3. In 2008, Germany’s per capita income in nominal terms was \$44,729 while its per capita income in PPP terms was \$35,539. Based on this information, if you were a German traveling in the United States, would U.S. products seem expensive or inexpensive relative to German products?

Chapter 18

Interest Rate Determination

Money is a critical component of a modern economy because it facilitates voluntary exchanges. What exactly money is and how it fulfills this role is not widely understood. This chapter defines money and explains how a country's central bank determines the amount of money available in an economy. It also shows how changes in the amount of money in a country influence two very important macroeconomic variables: the interest rate and the inflation rate.

18.1 Overview of Interest Rate Determination

LEARNING OBJECTIVE

1. Learn how a money market model, combining money supply and demand, influences the equilibrium interest rate in an economy.

This chapter describes how the supply of **money**¹ and the demand for money combine to affect the equilibrium interest rate in an economy. The model is called the **money market model**².

A country's **money supply**³ is mostly the amount of coin and currency in circulation and the total value of all checking accounts in banks. These two types of assets are the most liquid (i.e., most easily used to buy goods and services). The amount of money available to spend in an economy is mostly determined by the country's central bank. The bank can control the total amount of money in circulation by using several levers (or tools), the most important of which is the sale or purchase of U.S. government Treasury bonds. Central bank sales or purchases of Treasury bonds are called "**open market operations**"⁴.

Money demand⁵ refers to the demand by households, businesses, and the government, for highly liquid assets such as currency and checking account deposits. Money demand is affected by the desire to buy things soon, but it is also affected by the opportunity cost of holding money. The opportunity cost is the interest earnings one gives up on other assets to hold money.

If interest rates rise, households and businesses will likely allocate more of their asset holdings into interest-bearing accounts (these are usually not classified as money) and will hold less in the form of money. Since interest-bearing deposits are the primary source of funds used to lend in the financial sector, changes in total money demand affect the supply of loanable funds and in turn affect the interest rates on loans.

Money supply and money demand will equalize only at one average interest rate. Also, at this interest rate, the supply of loanable funds financial institutions wish to lend equals the amount that borrowers wish to borrow. Thus the equilibrium interest rate in the economy is the rate that equalizes money supply and money demand.

1. Any asset that serves as a unit of account and can be used as a medium of exchange for economic transactions. It is all assets that have a high degree of liquidity.
2. A model showing how the supply of money and the demand for money combine to affect the equilibrium interest rate in an economy.
3. Mostly, the amount of coin and currency in circulation and the total value of all checking accounts in banks.
4. Refers to central bank purchases or sales of U.S. government Treasury bonds or bills.
5. Refers to the demand by households, businesses, and the government, for highly liquid assets such as currency and checking account deposits.

Using the money market model, several important relationships between key economic variables are shown:

- When the money supply rises (falls), the equilibrium interest rate falls (rises).
- When the price level increases (decreases), the equilibrium interest rate rises (falls).
- When real GDP rises (falls), the equilibrium interest rate rises (falls).

Connections

The money market model connects with the foreign exchange (Forex) market because the interest rate in the economy, which is determined in the money market, determines the rate of return on domestic assets. In the Forex market, interest rates are given exogenously, which means they are determined through some process not specified in the model. However, that process of interest rate determination is described in the money market. Economists will sometimes say that once the money market model and Forex model are combined, interest rates have been “endogenized.” In other words, interest rates are now conceived as being determined by more fundamental factors (gross domestic product [GDP] and money supply) that are not given as exogenous.

The money market model also connects with the goods market model in that GDP, which is determined in the goods market, influences money demand and hence the interest rate in the money market model.

KEY TAKEAWAY

- The key results from the money market model are the following:
 - When the money supply rises (falls), the equilibrium interest rate falls (rises).
 - When the price level increases (decreases), the equilibrium interest rate rises (falls).
 - When real gross domestic product (GDP) rises (falls), the equilibrium interest rate rises (falls).

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term describing what is mostly composed of coin and currency in circulation and checking account deposits in a country.
 - b. The term describing the amount of money that households, businesses, and government want to hold or have available.
 - c. Of *increase, decrease, or stay the same*, this happens to the interest rate when the money supply falls.
 - d. Of *increase, decrease, or stay the same*, this happens to the interest rate when the domestic price level falls.
 - e. Of *increase, decrease, or stay the same*, this happens to the interest rate when real GDP falls.

18.2 Some Preliminaries

LEARNING OBJECTIVES

1. Recognize how casual uses of the term *money* differ from the more formal definition used in the money market model.
2. Learn how to interpret *the equilibrium interest rate* in a world in which there are many different interest rates applied and different types of loans and deposits.

There are several sources of confusion that can affect complete understanding of this basic model.

The first source of confusion concerns the use of the term “money.” In casual conversation, money is sometimes used more narrowly and sometimes more broadly than the formal definition. For example, someone might say, “I want to be a doctor so I will make a lot of money.” In this case, the person is really referring to income, not money, *per se*. Since income is typically paid using money, the everyday substitution of the term *money* for income does make sense, but it can lead to confusion in interpreting the forthcoming model. In general, people use the term *money* whenever they want to refer to a country’s coin and currency and anything these items are used for in payment. However, our formal definition of *money* also includes items that are not coin and currency. Checking account deposits are an example of a type of money included in the formal definition but not more casually thought of as money. Thus pay attention to the definition and description below and be sure to recognize that one’s common conception of money may or may not overlap precisely with the formal definition.

A second source of confusion involves our usage of the term *interest rate*. The model that will be developed will derive an equilibrium interest rate for the economy. However, everyone knows that there are many interest rates in the economy, and each of these rates is different. There are different rates for your checking and savings account, different rates on a car loan and mortgage, different rates on credit cards and government bonds. Thus it is typical to wonder what interest rate we are talking about when we describe *the equilibrium interest rate*.

It is important to note that financial institutions make money (here I really should say “make a profit”) by lending to one group at a higher rate than it borrows. In other words, financial institutions accept deposits from one group of people

(savers) and lend it to another group of people (borrowers.) If they charge a higher interest rate on their loans than they do on deposits, the bank will make a profit.

This implies that, in general, interest rates on deposits to financial institutions are lower than interest rates on their loans. When we talk about the equilibrium interest rate in the forthcoming model, it will mostly apply to the interest rates on deposits rather than loans. However, we also have a small problem in interpretation since different deposits have different interest rates. Thus which interest rate are we really talking about?

The best way to interpret the equilibrium interest rate in the model is as a kind of average interest rate on deposits. At the end of this chapter, we will discuss economic changes that lead to an increase or decrease in the equilibrium interest rate. We should take these changes to mean several things. First, that average interest rates on deposits will rise. Now, some of these rates may rise and a few may fall, but there will be pressure for the average to increase. Second, since banks may be expected to maintain their rate of profit (if possible) when average deposit interest rates do increase, average interest rates on loans will also increase. Again, some loan rates may rise and some fall, but the market pressure will tend to push them upward.

The implication is that when the equilibrium interest rate changes we should expect most interest rates to move in the same direction. Thus the equilibrium interest rate really is referring to an average interest rate across the entire economy, for deposits and for loans.

KEY TAKEAWAYS

- The term *money* is used causally in a different ways than we define it in the model: here *money* is defined as total value of coin and currency in circulation and checking account deposits at a point in time.
- The equilibrium interest rate in the money market model should be interpreted as an average interest rate across the entire economy, for deposits and for loans.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *higher, lower, or the same*, this is how average interest rates on bank deposits compare with average interest rates on bank loans.
 - b. The term used to describe the amount of money a person earns as wages.
 - c. When a person is asked how much money he has, he typically doesn’t think to include the current balance in this type of bank deposit.
2. Since there are many different interest rates on many types of loans and deposits, how do we interpret the equilibrium interest rate in the model?

18.3 What Is Money?

LEARNING OBJECTIVE

1. Learn why money exists and what purpose it serves.

The money supply in a country refers to a stock of assets that can be readily used to purchase goods and services. An asset is anything that has value. Anything that has value could potentially be used in exchange for other goods, services, or assets. However, some assets are more easily exchangeable than other assets.

Examples of assets include currency, checking account balances, stocks, bonds, whole life insurance policies, real estate, and automobiles. Currency—dollar bills in the United States, pounds sterling in Britain, and pesos in Mexico—is an asset that is readily exchangeable for goods and services within its respective countries. In contrast, real estate is an asset that is very difficult to use to buy goods. For example, no grocery store would accept ownership of a few square feet of your house in exchange for your weekly groceries. The idea of this transaction is unimaginable. Yet these two extreme cases can help us understand the distinction we make between assets classified as money and those not considered money. Most textbook definitions of *money* begin by defining several of money's key features.

Money as a Unit of Account

One of the most important features of money is its application as a unit of account. In other words, we choose to measure the value of goods, services, and assets in terms of currency or money. In ancient societies, shells, shovels, hoes, knives, cattle, and grain were used as money. In these cases, it would have been common to define the value of an item in terms of how many shells, or knives, or cows, and so on the item exchanges for. The standard unit of account in a country is its currency: dollars in the United States, yen in Japan, and euros in the European Union.

Money as a Medium of Exchange

The key distinguishing feature of money, as compared with other nonmoney assets, is its role as a medium of exchange. Coin, and later currency, came into existence primarily to serve as a vehicle for the exchange of goods and services. Rather than hauling around items that you might hope to barter exchange for other goods you need, it is easier and more efficient to carry coin and currency to purchase goods.

However, in order for money to function in this role, it must have widespread acceptability. Anyone selling something you want must be willing to accept the coin or currency you have. Their willingness to accept will in turn depend on the expectation that they'll be able to use that coin later to buy the goods they want.

Other types of assets are often not acceptable as a medium of exchange. For example, if I own a \$1,000 U.S. savings bond, I am unlikely to be able to use the bond to purchase items in a store. Bonds can be traded at a bank or a bond market, where exchanges of this sort are common, but not anywhere else. Thus bonds do not function as a medium of exchange.

Liquidity⁶ is a term used to describe the distinction made here between bonds and currency. An asset is said to be liquid if it is readily exchangeable for goods and services. An asset is illiquid if it is not easily exchangeable. Thus coin and currency are very liquid assets, while bonds are more illiquid. Real estate is an example of a very illiquid asset since it could take a considerable amount of time to convert the ownership share of a home into a spendable form.

Money as a Store of Value

Perhaps the least important characteristic of money is an ability to serve as a store of value. This is less important because it does not distinguish money from other assets. All assets serve as a store of value. As an example, if I want to save some income from each paycheck so that I can go on a vacation next year, I need to hold that income in a form that will maintain its purchasing power. One simple way to hold it is by cashing my paycheck and putting currency into an envelope. That money accumulating in the envelope will be easily used to purchase plane tickets and a hotel room when I take my vacation next year. In this way, holding currency will allow me to store value over time. On the other hand, I could cash each paycheck and deposit some of the money I want to save into my online stock trading account. With these funds I can purchase stocks, another form of asset. Next year, I can sell the stocks and use the money to take my vacation. Thus stocks represent a store of value as well.

KEY TAKEAWAY

6. Refers to the degree to which an asset is quickly and easily exchangeable for goods and services. For example, currency is highly liquid, but real estate is not liquid.

- Money is any asset that serves as a unit of account and can be used as a medium of exchange for economic transactions. It is all assets that have a high degree of liquidity. Money also serves as a store of value, but it is not unique in this role.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The three characteristics that are used to define money.
 - b. This characteristic of money is shared by real estate assets.
 - c. This characteristic of money allows us to compare the values of different products.
 - d. Without this characteristic of money, individuals would be forced to trade by barter.

18.4 Money Supply Measures

LEARNING OBJECTIVE

1. Learn the various definitions of money supply and their approximate values in the U.S. economy.

In the United States, the **Federal Reserve Bank**⁷ (or “Federal Reserve,” and more informally, “the Fed”) reports several distinct measures of the aggregate money supply. The narrowest measure, M1, includes only the most liquid assets. Higher numbers following an “M” reflect broader measures of money that include less liquid assets. Below is a description of M1–M3. However, unless otherwise specified, all later references to the money supply will relate to the M1 definition.

Money Supply Measure “M1”

M1 consists of the most highly liquid assets. That is, M1 includes all forms of assets that are easily exchangeable as payment for goods and services. It consists of coin and currency in circulation, traveler’s checks, demand deposits, and other checkable deposits.

The first item in M1 is currency and coin in circulation. In the United States, “currency” refers to \$1, \$5, \$10, \$20, \$50, and \$100 bills. U.S. “coin,” meanwhile, refers to pennies, nickels, dimes, and quarters. “In circulation” means that it has to be outside of banks, in people’s wallets or purses and businesses’ cash registers. Once the currency or coin is deposited in a bank, it is no longer considered to be in circulation, thus it is no longer a part of the M1 money supply.

The second item of M1 is traveler’s checks. Traveler’s checks are like currency, except that they have a form of insurance tied to them. If a traveler’s check is lost or stolen, the issuer will reimburse you for the loss.

The third item in M1 is demand deposits or checking account balances in banks. These consist of money individuals and businesses have deposited into an account in which a check can be written to pay for goods and services. When a check is presented to the bank, it represents a demand for transfer of funds from the check writer to the agent receiving the check. Since the funds must be disbursed on demand, we also refer to these as demand deposits.

7. The U.S. central bank that controls money supply in the country.

The final category in M1 is labeled “other checkable deposits.” This consists of two items; NOW accounts and ATS accounts. NOW stands for “negotiable orders of withdrawal.” A NOW account is exactly like a checking account except for one thing: it can earn interest. Thus checking accounts without interest are demand deposits and those with interest are NOW accounts. ATS stands for “automatic transfer service.” ATS accounts are savings accounts (also called time deposits) with one special feature. They can be drawn automatically to cover overdrafts from one’s checking account. Thus if an individual has a checking account with “overdraft protection” tied to their savings account, then the savings account is an ATS account.

Table 18.1 "Components of U.S. M1 Money Supply, November 2009" shows the M1 money supply for the U.S. economy as of January 2005. Notice that the largest component of M1, just over half, is the coin and currency in circulation. Traveler’s checks are an insignificant share at \$7.5 billion. Demand deposits and other checkable deposits almost equally split the remaining shares of M1 at close to 25 percent each. The total value of the M1 money supply is \$1,688 billion, which is over 10 percent of annual U.S. GDP.

Table 18.1 Components of U.S. M1 Money Supply, November 2009

	Billions (\$)	Total M1 (%)
Currency in Circulation	859.1	51
Traveler’s Checks	5.1	< 1
Demand Deposits	439.0	26
Other Checkable Deposits	385.4	23
Total M1 Money Supply	1,688.7	100

Source: Federal Reserve Statistical Release, Money Stock Measures, January 14, 2010. <http://www.federalreserve.gov/releases/h6/Current>.

Money Supply Measure “M2”

M2 is a broader measure of money than M1. It includes all of M1, the most liquid assets, and a collection of additional assets that are slightly less liquid. These additional assets include savings accounts, money market deposit accounts, small time deposits (less than \$100,000) and retail money market mutual funds. Excluded are IRA and Keogh deposits in money market accounts. (These are excluded since

they are retirement funds and hence are unlikely to be used as payment for goods and services anytime soon.)

Money Supply Measure “M3”

M3 is an even broader definition of the money supply, including M2 and other assets even less liquid than M2. As the number gets larger (i.e., “1, 2, 3...”), the assets included become less and less liquid. The additional assets include large-denomination time deposits (amounts greater than \$100,000), balances in institutional money funds (these include pension funds deposits), responsible party (RP) liabilities issued by depository institutions (refers to repurchase agreements), and eurodollars held by U.S. residents at foreign branches of U.S. banks worldwide and all banking offices in Canada and the United Kingdom (eurodollars are any U.S. dollar deposits made in a depository institution outside the United States). M3 excludes assets held by depository institutions, the U.S. government, money funds, and foreign banks and official institutions.

The United States values of all three major money supply definitions are given in Table 18.2 "U.S. Money Supply Measures (in Billions of Dollars), November 2009". Note that the M1 definition of money is just under one-tenth of the value of the annual GDP in the United States. The M2 money supply is almost six times larger, indicating substantial deposits in savings and time deposits and money market funds. M3 was last reported by the U.S. Fed in February 2006. But at that time, it was almost 90 percent of the U.S. annual GDP.

Table 18.2 U.S. Money Supply Measures (in Billions of Dollars), November 2009

M1	1,688.7
M2	8,391.9
M3 (February 2006)	10,298.7

Source: Federal Reserve Statistical Release, Money Stock Measures, January 14, 2010. For the most recent figures, go to <http://www.federalreserve.gov/releases/h6/Current>. (M3 was last reported for February 2006.)

KEY TAKEAWAYS

- M1 consists of the most highly liquid assets, including coin and currency in circulation, traveler's checks, demand deposits, and other checkable deposits.
- M2 is a broader measure of money than M1. It includes all of M1, plus savings accounts, money market deposit accounts, small-time deposits, and retail money market mutual funds.
- M3 is an even broader definition of the money supply that includes M2 plus large-denomination time deposits, balances in institutional money funds, repurchase liabilities, and eurodollars held by U.S. residents at foreign branches of U.S. banks.
- In 2009, the U.S. M1 was at just over \$1.6 trillion, around 10 percent of the U.S. gross domestic product (GDP).

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of M1, M2, or M3, this measure of money is the most liquid.
 - b. Of M1, M2, and/or M3, this measure(s) of money includes checking account deposits.
 - c. Of M1, M2, and/or M3, this measure(s) of money includes savings account deposits.
 - d. Of M1, M2, and/or M3, this measure(s) of money includes coin and currency in circulation.
 - e. Of M1, M2, and/or M3, this measure(s) of money includes eurodollars held by U.S. residents at foreign branches of U.S. banks.

18.5 Controlling the Money Supply

LEARNING OBJECTIVE

1. Learn the mechanisms (or tools) the U.S. Federal Reserve Bank can use to control the U.S. money supply.

The size of the money stock in a country is primarily controlled by its central bank. In the United States, the central bank is the Federal Reserve Bank while the main group affecting the money supply is the Federal Open Market Committee (FOMC). This committee meets approximately every six weeks and is the body that determines monetary policy. There are twelve voting members, including the seven members of the Fed Board of Governors and five presidents drawn from the twelve Federal Reserve banks on a rotating basis. The current Chairman of the Board of Governors is **Ben Bernanke** (as of January 2010). Because Bernanke heads the group that controls the money supply of the largest economy in the world, and because the FOMC's actions can have immediate and dramatic effects on interest rates and hence the overall United States and international economic condition, he is perhaps the most economically influential person in the world today. As you'll read later, because of his importance, anything he says in public can have tremendous repercussions throughout the international marketplace.

The Fed has three main levers that can be applied to affect the money supply within the economy: (1) open market operations, (2) **reserve requirement**⁸ changes, and (3) changes in the **discount rate**⁹.

The Fed's First Lever: Open Market Operations

- 8. Requirement that a fraction of the bank's total deposits be held in reserve either in the form of coin and currency in its vault or as a deposit (reserve) held at the Federal Reserve Bank.
- 9. Interest rate charged on the loans that the Federal Reserve Bank lends on a short-term basis (usually overnight) to financial institutions.

The most common lever used by the Fed is open market operations. This refers to Fed purchases or sales of U.S. government Treasury bonds or bills. The “open market” refers to the secondary market for these types of bonds. (The market is called secondary because the government originally issued the bonds at some time in the past.)

When the Fed purchases bonds on the open market it will result in an increase in the money supply. If it sells bonds on the open market, it will result in a decrease in the money supply.

Here's why. A purchase of bonds means the Fed buys a U.S. government Treasury bond from one of its primary dealers. This includes one of twenty-three financial institutions authorized to conduct trades with the Fed. These dealers regularly trade government bonds on the secondary market and treat the Fed as one of their regular customers. It is worth highlighting that bonds sold on the secondary open market are bonds issued by the government months or years before and will not mature for several months or years in the future. Thus when the Fed purchases a bond from a primary dealer in the future, when that bond matures, the government would have to pay back the Fed, which is the new owner of that bond.

When the **open market operation (OMO)**¹⁰ purchase is made, the Fed will credit that dealer's reserve deposits with the sale price of the bond (e.g., \$1 million). The Fed will receive the IOU, or "I owe you" (i.e., bond certificate), in exchange. The money used by the Fed to purchase this bond does not need to come from somewhere. The Fed doesn't need gold, other deposits, or anything else to cover this payment. Instead, the payment is created out of thin air. An accounting notation is made to indicate that the bank selling the bond now has an extra \$1 million in its reserve account.

At this point, there is still no change in the money supply. However, because of the increase in its reserves, the dealer now has additional money to lend out somewhere else, perhaps to earn a greater rate of return. When the dealer does lend it, it will create a demand deposit account for the borrower and since a demand deposit is a part of the M1 money supply, money has now been created.

As shown in all introductory macroeconomics textbooks, the initial loan, once spent by the borrower, is ultimately deposited in checking accounts in other banks. These increases in deposits can in turn lead to further loans, subject to maintenance of the bank's deposit reserve requirements. Each new loan made creates additional demand deposits and hence leads to further increases in the M1 money supply. This is called the money multiplier process. Through this process, each \$1 million bond purchase by the Fed can lead to an increase in the overall money supply many times that level.

The opposite effect will occur if the Fed sells a bond in an OMO. In this case, the Fed receives payment from a dealer (as in our previous example) in exchange for a previously issued government bond. (It is important to remember that the Fed does not issue government bonds; government bonds are issued by the U.S. Treasury department. If the Fed were holding a mature government bond, the Treasury would be obligated to pay off the face value to the Fed, just as if it were a private business or bank.) The payment made by the dealer comes from its reserve assets. These reserves support the dealer's abilities to make loans and in turn to stimulate

10. Refers to Fed purchases or sales of U.S. government Treasury bonds or bills. Used as a way to control the money supply.

the money creation process. Now that its reserves are reduced, the dealer's ability to create demand deposits via loans is reduced and hence the money supply is also reduced accordingly.

A more detailed description of open market operations can be found at New York Federal Reserve Bank's Web site at <http://www.ny.frb.org/aboutthefed/fedpoint/fed32.html>.

The Fed's Second Lever: Reserve Requirement Changes

When the Fed lowers the reserve requirement on deposits, the money supply increases. When the Fed raises the reserve requirement on deposits, the money supply decreases.

The reserve requirement is a rule set by the Fed that must be satisfied by all depository institutions, including commercial banks, savings banks, thrift institutions, and credit unions. The rule requires that a fraction of the bank's total transactions deposits (e.g., this would include checking accounts but not certificates of deposit) be held as a reserve either in the form of coin and currency in its vault or as a deposit (reserve) held at the Fed. The current reserve requirement in the United States (as of December 2009) is 10 percent for deposits over \$55.2 million. (For smaller banks—that is, those with lower total deposits—the reserve requirement is lower.)

As discussed above, the reserve requirement affects the ability of the banking system to create additional demand deposits through the money creation process. For example, with a reserve requirement of 10 percent, Bank A, which receives a deposit of \$100, will be allowed to lend out \$90 of that deposit, holding back \$10 as a reserve. The \$90 loan will result in the creation of a \$90 demand deposit in the name of the borrower, and since this is a part of the money supply M1, it rises accordingly. When the borrower spends the \$90, a check will be drawn on Bank A's deposits and this \$90 will be transferred to another checking account, say, in Bank B. Since Bank B's deposits have now risen by \$90, it will be allowed to lend out \$81 tomorrow, holding back \$9 (10 percent) as a reserve. This \$81 will make its way to another bank, leading to another increase in deposits, allowing another increase in loans, and so on. The total amount of demand deposits (DD) created through this process is given by the formula

$$DD = \$100 + (.9)\$100 + (.9)(.9)\$100 + (.9)(.9)(.9)\$100 + \dots$$

This simplifies to

$$DD = \$100 / (1 - 0.9) = \$1,000$$

or

$$DD = \$100 / RR,$$

where RR refers to the reserve requirement.

This example shows that if the reserve requirement is 10 percent, the Fed could increase the money supply by \$1,000 by purchasing a \$100 Treasury bill (T-bill) in the open market. However, if the reserve requirement were 5 percent, a \$100 T-bill purchase would lead to a \$2,000 increase in the money supply.

However, the reserve requirement not only affects the Fed's ability to create new money but also allows the banking system to create more demand deposits (hence more money) out of the total deposits it now has. Thus if the Fed were to lower the reserve requirement to 5 percent, the banking system would be able to increase the volume of its loans considerably and it would lead to a substantial increase in the money supply.

Because small changes in the reserve requirement can have substantial effects on the money supply, the Fed does not use reserve requirement changes as a primary lever to adjust the money supply.

A more detailed description of open market operations can be found at New York Federal Reserve Bank Web site at <http://www.ny.frb.org/aboutthefed/fedpoint/fed45.html>.

The Fed's Third Lever: Discount Rate/Federal Funds Rate Changes

When the Fed lowers its target federal funds rate and discount rate, it signals an expanded money supply and lower overall interest rates.

When the Fed raises its target federal funds rate and discount rate, it signals a reduced money supply and higher overall interest rates.

In news stories immediately after the FOMC meets, one is likely to read that the Fed raised (or lowered) interest rates yesterday. For many who read this, it sounds as if the Fed "sets" the interest rates charged by banks. In actuality, the Fed only sets

one interest rate, and that is the discount rate. The rate that is announced every month is not the discount rate, but the federal funds rate. The **federal funds rate**¹¹ is the interest rate banks charge each other for short-term (usually overnight) loans. The Fed does not actually set the federal funds rate, but it does employ open market operations to target this rate at a desired level. Thus what is announced at the end of each FOMC meeting is the target federal funds rate.

The main reason banks make overnight loans to each other each day is to maintain their reserve requirements. Each day some banks may end up with excess reserves. Other banks may find themselves short of reserves. Those banks with excess reserves would prefer to loan out as much as possible at some rate of interest rather than earning nothing. Those banks short of reserves are required by law to raise up their reserves to the required level. Thus banks lend money to each other each night.

If there is excess demand for money overnight relative to supply, the Fed keeps the discount window open. The discount window refers to a policy by the Fed to lend money on a short-term basis (usually overnight) to financial institutions. The interest rate charged on these loans is called the discount rate. Before 2003, banks needed to demonstrate that they had exhausted all other options before coming to the discount window. After 2003, the Fed revised its policies and set a primary credit discount rate and a secondary credit discount rate. Primary credit rates are set 100 basis points (1 percent) above the federal funds rate and are available only to very sound, financially strong banks. Secondary credit rates are set 150 basis points above the federal funds target rate and are available to banks not eligible for primary credit. Although these loans are typically made overnight, they can be extended for longer periods and can be used for any purpose.

Before the changes in discount window policy in 2003, very few banks sought loans through the discount window. Hence, it was not a very effective lever in monetary policy.

However, the announcement of the federal funds target rate after each FOMC meeting does remain an important signal about the future course of Fed monetary policy. If the FOMC announces a lower target federal funds rate, one should expect expanded money supply, perhaps achieved through open market operations. If the FOMC announces a higher target rate, one should prepare for a more **contractionary monetary policy**¹² to follow.

11. The interest rate banks charge each other for short-term (usually overnight) loans.

12. A decrease in the money supply in a country.

A more detailed description of the discount window can be found on the New York Federal Reserve Bank Web site at <http://www.ny.frb.org/aboutthefed/fedpoint/>

[fed18.html](#). For more information about federal funds, go to <http://www.ny.frb.org/aboutthefed/fedpoint/fed15.html>.

KEY TAKEAWAYS

- When the Federal Reserve Bank (a.k.a. “Federal Reserve,” or more informally, “the Fed”) purchases bonds on the open market it will result in an increase in the U.S. money supply. If it sells bonds in the open market, it will result in a decrease in the money supply.
- When the Fed lowers the reserve requirement on deposits, the U.S. money supply increases. When the Fed raises the reserve requirement on deposits, the money supply decreases.
- When the Fed lowers its target federal funds rate and discount rate, it signals an expanded U.S. money supply and lower overall interest rates.
- When the Fed raises its target federal funds rate and discount rate, it signals a reduced U.S. money supply and higher overall interest rates.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or no change*, the effect on the money supply if the central bank sells government bonds.
 - b. Of *increase, decrease, or no change*, the effect on the money supply if the central bank lowers the reserve requirement.
 - c. Of *increase, decrease, or no change*, the effect on the money supply if the central bank lowers the discount rate.
 - d. The name given to the interest rate charged by the Federal Reserve Bank on loans it provides to commercial banks
 - e. The name given to the interest rate charged by commercial banks on overnight loans made to other banks.

18.6 Money Demand

LEARNING OBJECTIVE

1. Learn the determinants of money demand in an economy.

The demand for money represents the desire of households and businesses to hold assets in a form that can be easily exchanged for goods and services. Spendability (or liquidity) is the key aspect of money that distinguishes it from other types of assets. For this reason, the demand for money is sometimes called the demand for liquidity.

The demand for money is often broken into two distinct categories: the transactions demand and the speculative demand.

Transactions Demand for Money

The primary reason people hold money is because they expect to use it to buy something sometime soon. In other words, people expect to make transactions for goods or services. How much money a person holds onto should probably depend on the value of the transactions that are anticipated. Thus a person on vacation might demand more money than on a typical day. Wealthier people might also demand more money because their average daily expenditures are higher than the average person's.

However, in this section we are interested not so much in an individual's demand for money but rather in what determines the aggregate, economy-wide demand for money. Extrapolating from the individual to the group, we could conclude that the total value of all transactions in the economy during a period would influence the aggregate transactions demand for money. Gross domestic product (GDP), the value of all goods and services produced during the year, will influence the aggregate value of all transactions since all GDP produced will be purchased by someone during the year. GDP may underestimate the demand for money, though, since people will also need money to buy used goods, intermediate goods, and assets. Nonetheless, changes in GDP are very likely to affect transactions demand.

Anytime GDP rises, there will be a demand for more money to make the transactions necessary to buy the extra GDP. If GDP falls, then people demand less money for transactions.

The GDP that matters here is nominal GDP, meaning GDP measured in terms of the prices that currently prevail (GDP at current prices). Economists often break up GDP into a nominal component and a real component, where real GDP corresponds to a quantity of goods and services produced after eliminating any price level changes that have occurred since the price level base year. To convert nominal to real GDP, simply divide nominal GDP by the current U.S. price level ($P_{\$}$); thus

$$\text{real GDP} = \text{nominal GDP}/P_{\$}.$$

If we use the variable $Y_{\$}$ to represent real U.S. GDP and rearrange the equation, we can get

$$\text{nominal GDP} = P_{\$} Y_{\$}.$$

By rewriting in this way we can now indicate that since the transactions demand for money rises with an increase in nominal GDP, it will also rise with either an increase in the general price level or an increase in real GDP.

Thus if the amount of goods and services produced in the economy rises while the prices of all products remain the same, then total GDP will rise and people will demand more money to make the additional transactions. On the other hand, if the average prices of goods and services produced in the economy rise, then even if the economy produces no additional products, people will still demand more money to purchase the higher valued GDP, hence the demand for money to make transactions will rise.

Speculative Demand for Money

The second type of money demand arises by considering the opportunity cost of holding money. Recall that holding money is just one of many ways to hold value or wealth. Alternative opportunities include holding wealth in the form of savings deposits, certificate of deposits, mutual funds, stock, or even real estate. For many of these alternative assets interest payments, or at least a positive rate of return, may be obtained. Most assets considered money, such as coin and currency and most checking account deposits, do not pay any interest. If one does hold money in the form of a negotiable order of withdrawal (NOW) account, a checking account with interest, the interest earned on that deposit will almost surely be less than on a savings deposit at the same institution.

Thus to hold money implies giving up the opportunity of holding other assets that pay interest. The interest one gives up is the opportunity cost of holding money.

Since holding money is costly—that is, there is an opportunity cost—people's demand for money should be affected by changes in its cost. Since the interest rate on each person's next best opportunity may differ across money holders, we can use the average interest rate ($i_{\$}$) in the economy as a proxy for the opportunity cost. It is likely that as average interest rates rise, the opportunity cost of holding money for all money holders will also rise, and vice versa. And as the cost of holding money rises, people should demand less money.

The intuition is straightforward, especially if we exaggerate the story. Suppose interest rates on time deposits suddenly increased to 50 percent per year (from a very low base). Such a high interest rate would undoubtedly lead individuals and businesses to reduce the amount of cash they hold, preferring instead to shift it into the high-interest-yielding time deposits. The same relationship is quite likely to hold even for much smaller changes in interest rates. This implies that as interest rates rise (fall), the demand for money will fall (rise). The speculative demand for money, then, simply relates to component of the money demand related to interest rate effects.

KEY TAKEAWAYS

- Anytime the gross domestic product (GDP) rises, there will be a demand for more money to make the transactions necessary to buy the extra GDP. If GDP falls, then people demand less money for transactions.
- The interest one gives up is the opportunity cost of holding money.
- As interest rates rise (fall), the demand for money will fall (rise).

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or no change*, the effect on the transactions demand for money when interest rates fall.
 - b. Of *increase, decrease, or no change*, the effect on the transactions demand for money when GDP falls.
 - c. Of *increase, decrease, or no change*, the effect on the speculative demand for money when GDP falls.
 - d. Of *increase, decrease, or no change*, the effect on the speculative demand for money when interest rates fall.

18.7 Money Functions and Equilibrium

LEARNING OBJECTIVE

1. Define real money demand and supply functions, graph them relative to the interest rate, and use them to define the equilibrium interest rate in an economy.

Demand

A money demand function displays the influence that some aggregate economic variables will have on the aggregate demand for money. The above discussion indicates that money demand will depend positively on the level of real gross domestic product (GDP) and the price level due to the demand for transactions. Money demand will depend negatively on average interest rates due to speculative concerns. We can depict these relationships by simply using the following functional representation:

$$M^D = f(P_{\$}^{+}, Y_{\$}^{+}, i_{\$}^{-}).$$

Here M^D is the aggregate, economy-wide money demand, $P_{\$}$ is the current U.S. price level, $Y_{\$}$ is the United States' real GDP, and $i_{\$}$ is the average U.S. interest rate. The f stands for “function.” The f is not a variable or parameter value; it simply means that some function exists that would map values for the right-side variables, contained within the brackets, into the left-side variable. The “+” symbol above the price level and GDP levels means that there is a positive relationship between changes in that variable and changes in money demand. For example, an increase (decrease) in $P_{\$}$ would cause an increase (decrease) in M^D . A “-” symbol above the interest rate indicates that changes in $i_{\$}$ in one direction will cause money demand to change in the opposite direction.

For historical reasons, the money demand function is often transformed into a real money demand function as follows. First, rewrite the function on the right side to get

$$M^D = P_{\$}^{+} L(Y_{\$}^{+}, i_{\$}^{-}).$$

In this version, the price level ($P_{\$}$) is brought outside the function $f(\cdot)$ and multiplied to a new function labeled $L(\cdot)$, called the “liquidity function.” Note that $L(\cdot)$ is different from $f(\cdot)$ since it contains only $Y_{\$}$ and $i_{\$}$ as variables. Since $P_{\$}$ is multiplied to $L(\cdot)$ it will maintain the positive relationship to M^D and thus is perfectly consistent with the previous specification. Finally, by moving the price level variable to the left side, we can write out the general form of the real money demand function as

$$\frac{M^D}{P_{\$}} = L(Y_{\$}, i_{\$}).$$

This states that real money demand ($M^D/P_{\$}$) is positively related to changes in real GDP ($Y_{\$}$) and the average interest rate ($i_{\$}$) according to the liquidity function. We can also say that the liquidity function represents the real demand for money in the economy—that is, the liquidity function is equivalent to real money demand.

Finally, simply for intuition’s sake, any real variable represents the purchasing power of the variable in terms of prices that prevailed in the base year of the price index. Thus real money demand can be thought of as the purchasing power of money demanded in terms of base year prices.

Supply

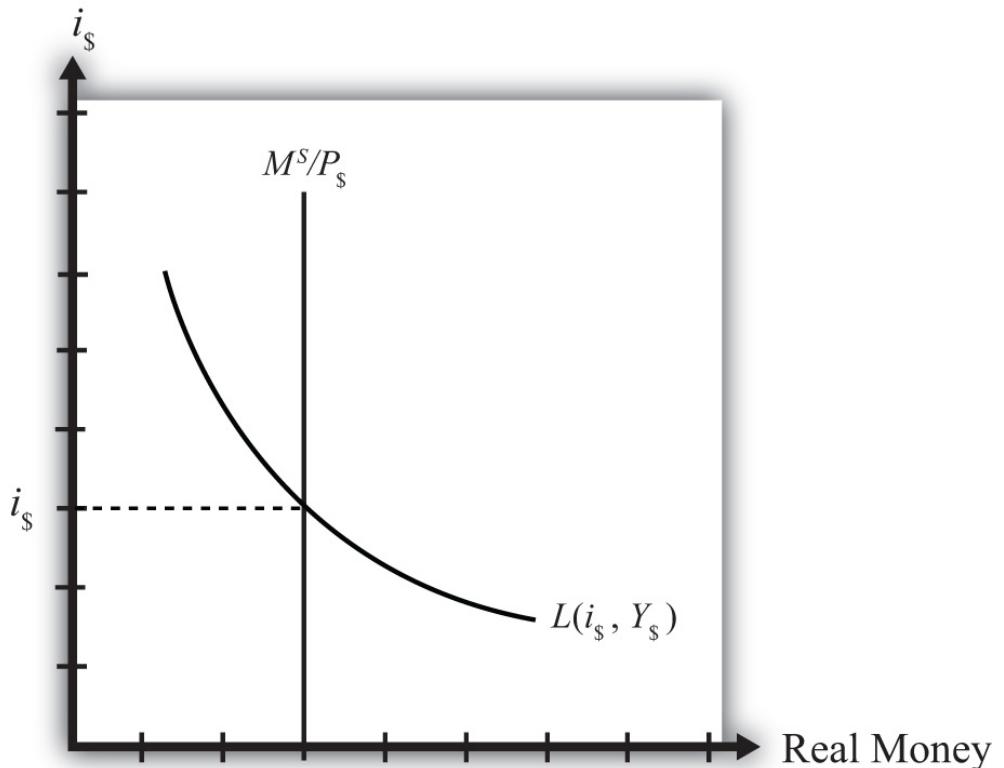
Money supply is much easier to describe because we imagine that the level of money balances available in an economy is simply set by the actions of the central bank. For this reason, it will not depend on other aggregate variables such as the interest rate, and thus we need no function to describe it.

We will use the parameter $M_{\S to represent the nominal U.S. money supply and assume that the Federal Reserve Bank (or simply “the Fed”), using its three levers, can set this variable wherever it chooses. To represent real money supply, however, we will need to convert by dividing by the price level. Hence let $\frac{M_{\$}^S}{P_{\$}}$ represent the real money supply in terms of prices that prevailed in the base year.

Equilibrium

The equilibrium interest rate is determined at the level that will equalize real money supply with real money demand. We can depict the equilibrium by graphing the money supply and demand functions on the following diagram.

Figure 18.1 The Money Market



The functions are drawn in [Figure 18.1 "The Money Market"](#) with real money, both supply and demand, plotted along the horizontal axis and the interest rate plotted along the vertical axis.

Real money supply, $\frac{M^S}{P_{\$}}$, is drawn as a vertical line at the level of money balances, measured best by M1. It is vertical because changes in the interest rate will not affect the money supply in the economy.

Real money demand—that is, the liquidity function $L(i_{\$}, Y_{\$})$ —is a downward sloping line in $i_{\$}$ reflecting the speculative demand for money. In other words, there is a negative relationship presumed to prevail between the interest rate and real money demand.

Where the two lines cross determines the equilibrium interest rate in the economy ($i_{\$}$) since this is the only interest rate that will equalize real money supply with real money demand.

KEY TAKEAWAYS

- Real money demand is positively related to changes in real gross domestic product (GDP) and the average interest rate.
- Real money supply is independent of the average interest rate and is assumed to be determined by the central bank.
- The intersection of the real money supply function and the real money demand function determines the equilibrium interest rate in the economy.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *positive, negative, or no effect*, this is the relationship between the interest rate and real money demand.
 - b. Of *positive, negative, or no effect*, this is the relationship between real GDP and real money demand.
 - c. Of *positive, negative, or no effect*, this is the relationship between the price level and nominal money demand.
 - d. Of *positive, negative, or no effect*, this is the relationship between the interest rate and real money supply.
 - e. Of *positive, negative, or no effect*, this is the relationship between real GDP and real money supply.
 - f. Of *positive, negative, or no effect*, this is the relationship between the price level and real money supply.
 - g. The endogenous variable (in the money market model) whose value is determined at the intersection of the real money supply curve and the real money demand curve.

18.8 Money Market Equilibrium Stories

LEARNING OBJECTIVE

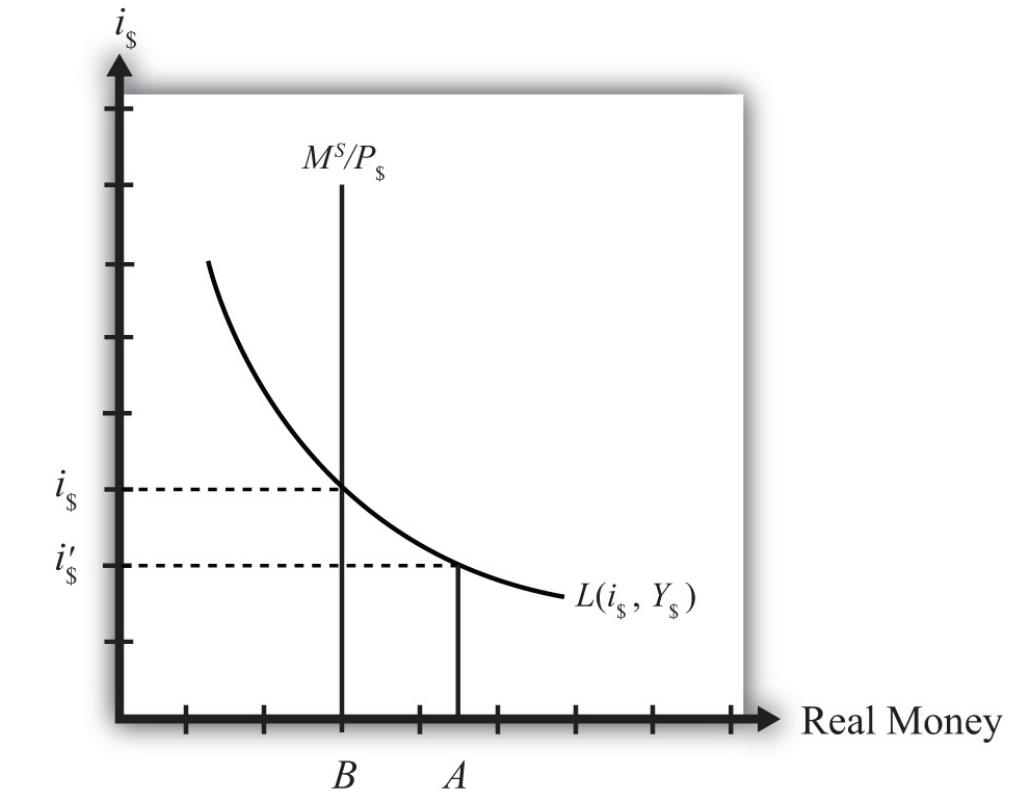
1. Learn the equilibrium stories in the money market that describe how the interest rate adjusts when it is not at its equilibrium value.

Any equilibrium in economics has an associated behavioral story to explain the forces that will move the endogenous variable to the equilibrium value. In the money market model, the endogenous variable is the interest rate. This is the variable that will change to achieve the equilibrium. Variables that do not change in the adjustment to the equilibrium are the exogenous variables. In this model, the exogenous variables are $P_{\$}$, $Y_{\$}$, and $M_{\S . Changes in the exogenous variables are necessary to cause an adjustment to a new equilibrium. However, in telling an equilibrium story, it is typical to simply assume that the endogenous variable is not at the equilibrium (for some unstated reason) and then to explain how and why the variable will adjust to the equilibrium value.

Interest Rate Too Low

Suppose that for some reason the actual interest rate, $i'_{\$}$ lies below the equilibrium interest rate ($i_{\$}$) as shown in [Figure 18.2 "Adjustment to Equilibrium: Interest Rate Too Low"](#). At $i'_{\$}$, real money demand is given by the value A along the horizontal axis, while real money supply is given by the value B. Since A is to the right of B, real demand for money exceeds the real money supply. This means that people and businesses wish to hold more assets in a liquid, spendable form rather than holding assets in a less liquid form, such as in a savings account. This excess demand for money will cause households and businesses to convert assets from less liquid accounts into checking accounts or cash in their pockets. A typical transaction would involve a person who withdraws money from a savings account to hold cash in his wallet.

Figure 18.2 Adjustment to Equilibrium: Interest Rate Too Low



The savings account balance is not considered a part of the M1 money supply; however, the currency the person puts into his wallet is a part of the money supply. Millions of conversions such as this will be the behavioral response to an interest rate that is below equilibrium. As a result, the financial sector will experience a decrease in time deposit balances, which in turn will reduce their capacity to make loans. In other words, withdrawals from savings and other type of nonmoney accounts will reduce the total pool of funds available to be loaned by the financial sector. With fewer funds to lend and the same demand for loans, banks will respond by raising interest rates. Higher interest rates will reduce the demand for loans helping to equalize supply and demand for loans. Finally, as interest rates rise, money demand falls until it equals with the actual money supply. Through this mechanism average interest rates will rise, whenever money demand exceeds money supply.

Interest Rate Too High

If the actual interest rate is higher than the equilibrium rate, for some unspecified reason, then the opposite adjustment will occur. In this case, real money supply will exceed real money demand, meaning that the amount of assets or wealth people

and businesses are holding in a liquid, spendable form is greater than the amount they would like to hold. The behavioral response would be to convert assets from money into interest-bearing nonmoney deposits. A typical transaction would be if a person deposits some of the cash in his wallet into his savings account. This transaction would reduce money holdings since currency in circulation is reduced, but will increase the amount of funds available to loan out by the banks. The increase in loanable funds, in the face of constant demand for loans, will inspire banks to lower interest rates to stimulate the demand for loans. However, as interest rates fall, the demand for money will rise until it equalizes again with money supply. Through this mechanism average interest rates will fall whenever money supply exceeds money demand.

KEY TAKEAWAYS

- If the actual interest rate is lower than the equilibrium rate, the amount of assets people are holding in a liquid form is less than the amount they would like to hold. They respond by converting assets from interest-bearing nonmoney deposits into money. The decrease in loanable funds will cause banks to raise interest rates. Interest rates rise until money supply equals money demand.
- If the actual interest rate is higher than the equilibrium rate, the amount of assets people are holding in a liquid form is greater than the amount they would like to be holding. They respond by converting assets from money into interest-bearing nonmoney deposits. The increase in loanable funds will cause banks to lower interest rates. Interest rates fall until money supply equals money demand.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the effect on the average interest rate when real money supply exceeds real money demand.
 - b. Of *increase, decrease, or stay the same*, the effect on the average interest rate when real money demand is less than real money supply.
 - c. Of *increase, decrease, or stay the same*, the effect on the average interest rate when real money demand exceeds real money supply.
 - d. Of *increase, decrease, or stay the same*, the effect on the average interest rate when households and businesses wish to convert assets from interest-bearing nonmoney deposits into money.
 - e. Of *increase, decrease, or stay the same*, the effect on the average interest rate when households and businesses wish to convert assets from money into interest-bearing nonmoney deposits.

18.9 Effects of a Money Supply Increase

LEARNING OBJECTIVE

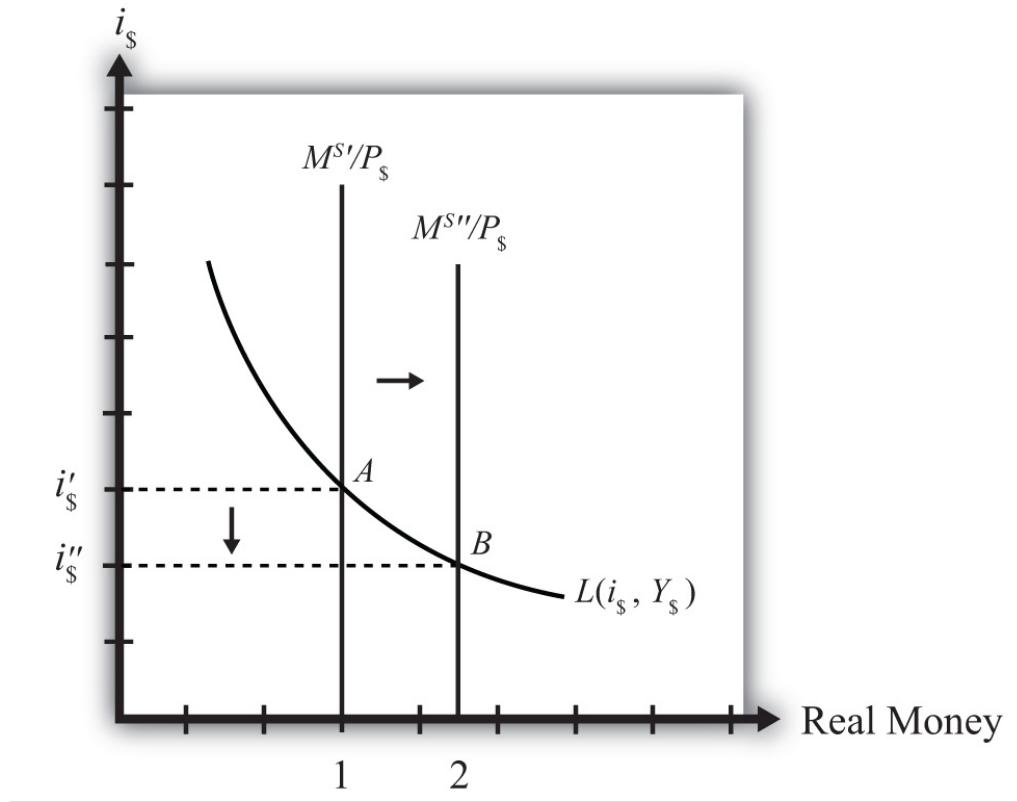
1. Learn how a change in the money supply affects the equilibrium interest rate.

Expansionary monetary policy¹³ refers to any policy initiative by a country's central bank to raise (or expand) its money supply. This can be accomplished with open market purchases of government bonds, with a decrease in the reserve requirement, or with an announced decrease in the discount rate. In most growing economies the money supply is expanded regularly to keep up with the expansion of gross domestic product (GDP). In this dynamic context, expansionary monetary policy can mean an increase in the *rate of growth* of the money supply, rather than a mere increase in money. However, the money market model is a nondynamic (or static) model, so we cannot easily incorporate money supply growth rates. Nonetheless, we can project the results from this static model to the dynamic world without much loss of relevance. (In contrast, any decrease in the money supply or decrease in the growth rate of the money supply is referred to as *contractionary monetary policy*.)

Suppose the money market is originally in equilibrium in [Figure 18.3 "Effects of a Money Supply Increase"](#) at point A with real money supply $M^S/P_{\$}$ and interest rate $i_{\$}'$ when the money supply increases, *ceteris paribus*. The *ceteris paribus* assumption means we assume that all other exogenous variables in the model remain fixed at their original levels. In this exercise, it means that real GDP ($Y_{\$}$) and the price level ($P_{\$}$) remain fixed. An increase in the money supply (M^S) causes an increase in the real money supply ($M^S/P_{\$}$) since $P_{\$}$ remains constant. In the diagram, this is shown as a rightward shift from $M^S/P_{\$}$ to $M^S''/P_{\$}$. At the original interest rate, real money supply has risen to level 2 along the horizontal axis while real money demand remains at level 1. This means that money supply exceeds money demand, and the actual interest rate is higher than the equilibrium rate. Adjustment to the lower interest rate will follow the “interest rate too high” equilibrium story.

13. An increase in the money supply in a country.

Figure 18.3 Effects of a Money Supply Increase



The final equilibrium will occur at point *B* on the diagram. The real money supply will have risen from level 1 to 2 while the equilibrium interest rate has fallen from $i_{\$}'$ to $i_{\$}''$. Thus expansionary monetary policy (i.e., an increase in the money supply) will cause a decrease in average interest rates in an economy. In contrast, contractionary monetary policy (a decrease in the money supply) will cause an increase in average interest rates in an economy.

Note this result represents the short-run effect of a money supply increase. The short run is the time before the money supply can affect the price level in the economy. In [Chapter 18 "Interest Rate Determination"](#), [Section 18.14 "Money Supply and Long-Run Prices"](#), we consider the long-run effects of a money supply increase. In the long run, money supply changes can affect the price level in the economy. In the previous exercise, since the price level remained fixed (i.e., subject to the *ceteris paribus* assumption) when the money supply was increased, this exercise provides the short-run result.

KEY TAKEAWAY

- An increase (decrease) in the money supply, ceteris paribus, will cause a decrease (increase) in average interest rates in an economy.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Term often used to describe the type of monetary policy that results in a reduction of the money supply.
 - b. Term often used to describe the type of monetary policy that results in an increase in the money supply.
 - c. Of *increase, decrease, or stay the same*, the effect on the equilibrium interest rate when the nominal money supply increases, ceteris paribus.
 - d. Of *increase, decrease, or stay the same*, the effect on the equilibrium interest rate when the nominal money supply decreases, ceteris paribus.
 - e. Term for the time period before price level changes occur in the money market model.

18.10 Effect of a Price Level Increase (Inflation) on Interest Rates

LEARNING OBJECTIVE

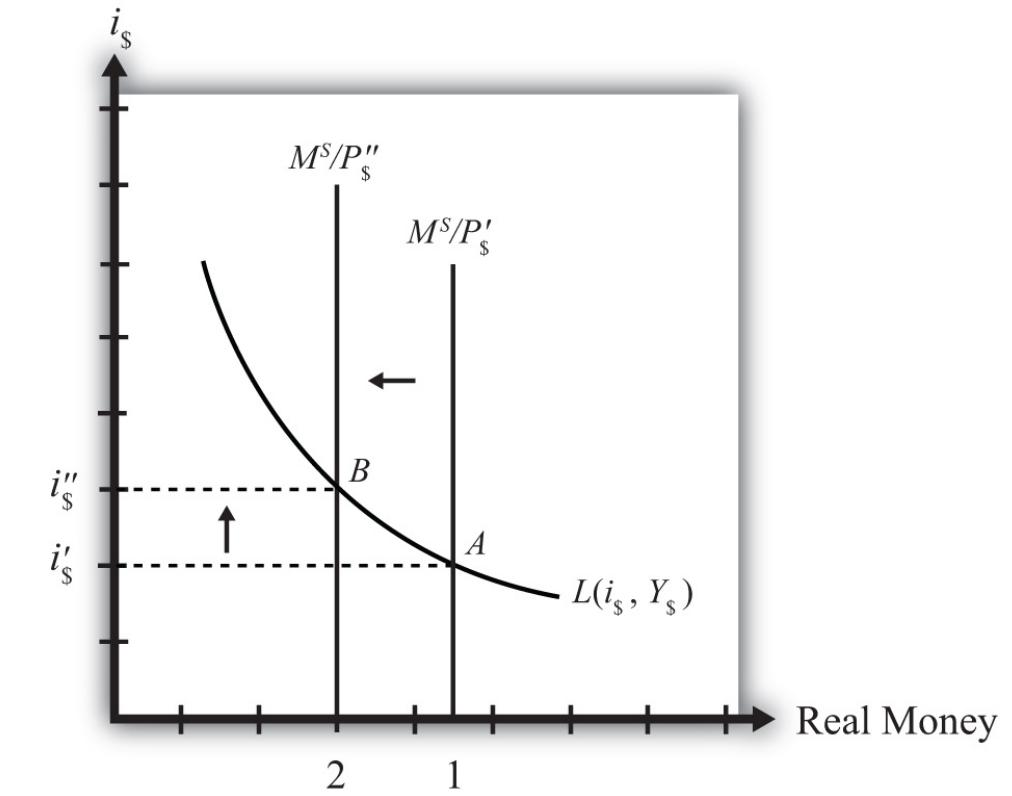
1. Learn how a change in the price level affects the equilibrium interest rate.

Now let's consider the effects of a price level increase in the money market. When the price level rises in an economy, the average price of all goods and services sold is increasing. Inflation is calculated as the percentage increase in a country's price level over some period, usually a year. This means that in the period during which the price level increases, inflation is occurring. Thus studying the effects of a price level increase is the same as studying the effects of inflation.

Inflation can arise for several reasons that will be discussed later in this chapter. For now, we will imagine that the price level increases for some unspecified reason and consider the consequences.

Suppose the money market is originally in equilibrium at point A in [Figure 18.4 "Effects of a Price Level Increase"](#) with real money supply $M^S/P\$'$ and interest rate $i\$'$. Suppose the price level increases, ceteris paribus. Again, the ceteris paribus assumption means that we assume all other exogenous variables in the model remain fixed at their original levels. In this exercise, it means that the money supply (M^S) and real GDP ($Y\$$) remain fixed. An increase in the price level ($P\$$) causes a decrease in the real money supply ($M^S/P\$$) since M^S remains constant. In the adjoining diagram, this is shown as a shift from $M^S/P\$'$ to $M^S/P\$''$. At the original interest rate, $i\$'$, the real money supply has fallen to level 2 along the horizontal axis, while real money demand remains at level 1. This means that money demand exceeds money supply and the actual interest rate is lower than the new equilibrium rate. Adjustment to the higher interest rate will follow the “interest rate too low” equilibrium story.

Figure 18.4 Effects of a Price Level Increase



More intuition concerning these effects arises if one recalls that price level increases will increase the transactions demand for money. In this version, nominal money demand will exceed nominal money supply and set off the same adjustment process described in the previous paragraph.

The final equilibrium will occur at point *B* on the diagram. The real money supply will have fallen from level 1 to level 2 while the equilibrium interest rate has risen from $i_{\$}'$ to $i_{\$}''$. Thus an increase in the price level (i.e., inflation) will cause an increase in average interest rates in an economy. In contrast, a decrease in the price level (deflation) will cause a decrease in average interest rates in an economy.

KEY TAKEAWAY

- An increase in the price level (i.e., inflation), ceteris paribus, will cause an increase in average interest rates in an economy. In contrast, a decrease in the price level (deflation), ceteris paribus, will cause a decrease in average interest rates in an economy.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a percentage increase in a country’s price level over a period of time.
 - b. Of *increase, decrease, or stay the same*, the effect on the equilibrium interest rate when the domestic price level decreases, *ceteris paribus*.
 - c. Of *increase, decrease, or stay the same*, the effect on the equilibrium interest rate when the domestic price level increases, *ceteris paribus*.

18.11 Effect of a Real GDP Increase (Economic Growth) on Interest Rates

LEARNING OBJECTIVE

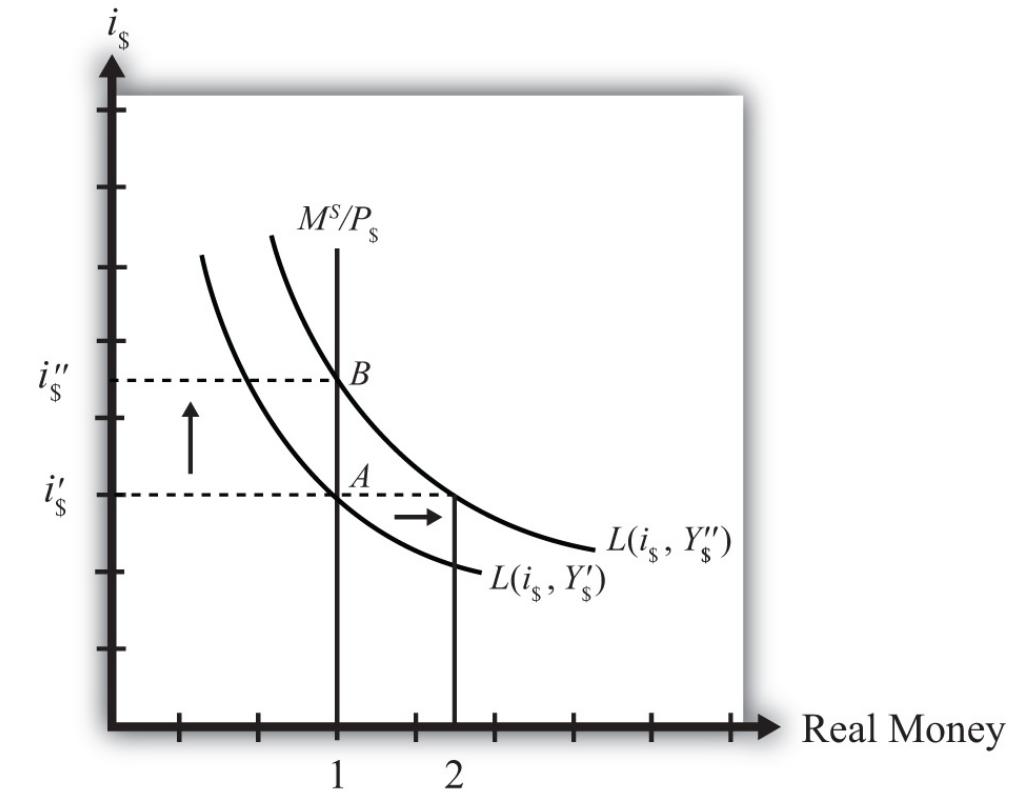
1. Learn how a change in real GDP affects the equilibrium interest rate.

Finally, let's consider the effects of an increase in real gross domestic product (GDP). Such an increase represents economic growth. Thus the study of the effects of a real GDP increase is the same as asking how economic growth will affect interest rates.

GDP may increase for a variety of reasons, which are discussed in subsequent chapters. For now, we will imagine that GDP increases for some unspecified reason and consider the consequences of such a change in the money market.

Suppose the money market is originally in equilibrium at point A in [Figure 18.5 "Effects of an Increase in Real GDP"](#) with real money supply $M^S/P\$$ and interest rate $i\$'$. Suppose real GDP ($Y\$$) increases, ceteris paribus. Again, the ceteris paribus assumption means that we assume all other exogenous variables in the model remain fixed at their original levels. In this exercise, it means that the money supply (M^S) and the price level ($P\$$) remain fixed. An increase in GDP will raise the demand for money because people will need more money to make the transactions necessary to purchase the new GDP. In other words, real money demand rises due to the transactions demand effect. This increase is reflected in the rightward shift of the real money demand function from $L(i\$, Y\$')$ to $L(i\$, Y\$'')$.

Figure 18.5 Effects of an Increase in Real GDP



At the original interest rate, $i'_{\$}$, real money demand has increased to level 2 along the horizontal axis while real money supply remains at level 1. This means that real money demand exceeds real money supply and the current interest rate is lower than the equilibrium rate. Adjustment to the higher interest rate will follow the “interest rate too low” equilibrium story.

The final equilibrium will occur at point *B* on the diagram. As the interest rate rises from $i'_{\$}$ to $i''_{\$}$, real money demand will have fallen from level 2 to level 1. Thus an increase in real GDP (i.e., economic growth) will cause an increase in average interest rates in an economy. In contrast, a decrease in real GDP (a recession) will cause a decrease in average interest rates in an economy.

KEY TAKEAWAY

- An increase in real gross domestic product (i.e., economic growth), *ceteris paribus*, will cause an increase in average interest rates in an economy. In contrast, a decrease in real GDP (a recession), *ceteris paribus*, will cause a decrease in average interest rates in an economy.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe a percentage increase in real GDP over a period of time.
 - b. Of *increase, decrease, or stay the same*, the effect on the equilibrium interest rate when real GDP decreases, *ceteris paribus*.
 - c. Of *increase, decrease, or stay the same*, the effect on the equilibrium interest rate when real GDP increases, *ceteris paribus*.

18.12 Integrating the Money Market and the Foreign Exchange Markets

LEARNING OBJECTIVE

1. Integrate the money market with the foreign exchange market and highlight the interactions that exist between the two.

In this section, we will integrate the money market with the foreign exchange market to demonstrate the interactions that exist between the two. First, let's review.

In the money market, the endogenous variable is the interest rate ($i_{\$}$). This is the variable that is determined in equilibrium in the model. The exogenous variables are the money supply (M^S), the price level ($P_{\$}$), and the level of real gross domestic product (GDP) (Y). These variables are determined outside the money market and treated as known values. Their values determine the supply and demand for money and affect the equilibrium value of the interest rate.

In the foreign exchange (Forex) market, the endogenous variable is the exchange rate, $E_{\$/\text{£}}$. The exogenous variables are the domestic interest rate ($i_{\$}$), the foreign interest rate ($i_{\text{£}}$), and the expected exchange rate ($E_{\$/\text{£}}^e$). Their values determine the domestic and foreign rates of return and affect the equilibrium value of the exchange rate.

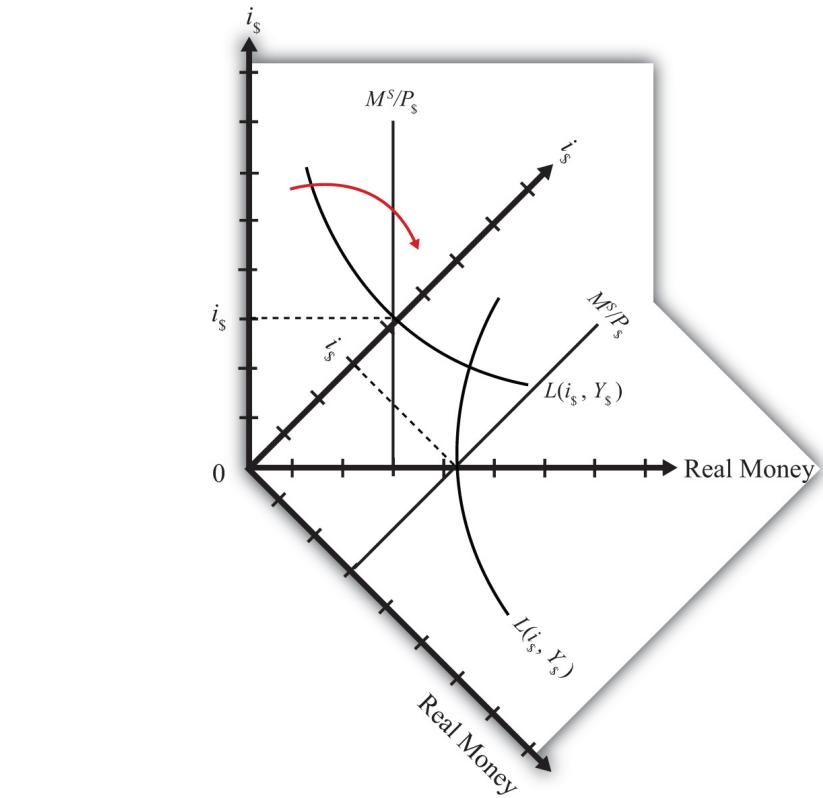
The linkage between the two markets arises because the domestic interest rate is the endogenous variable in the money market and an exogenous variable in the Forex market. Thus when considering the Forex, when we say the interest rate is determined outside of the Forex market, we know where it is determined: it is determined in the U.S. money market as the interest rate that satisfies real supply and demand for money.

Linking the Diagrams

We can keep track of the interactions between these two markets using a simple graphical technique. We begin with the money market diagram as developed in [Chapter 18 "Interest Rate Determination"](#), [Section 18.7 "Money Functions and Equilibrium"](#). The trick is to rotate the diagram ninety degrees in a clockwise

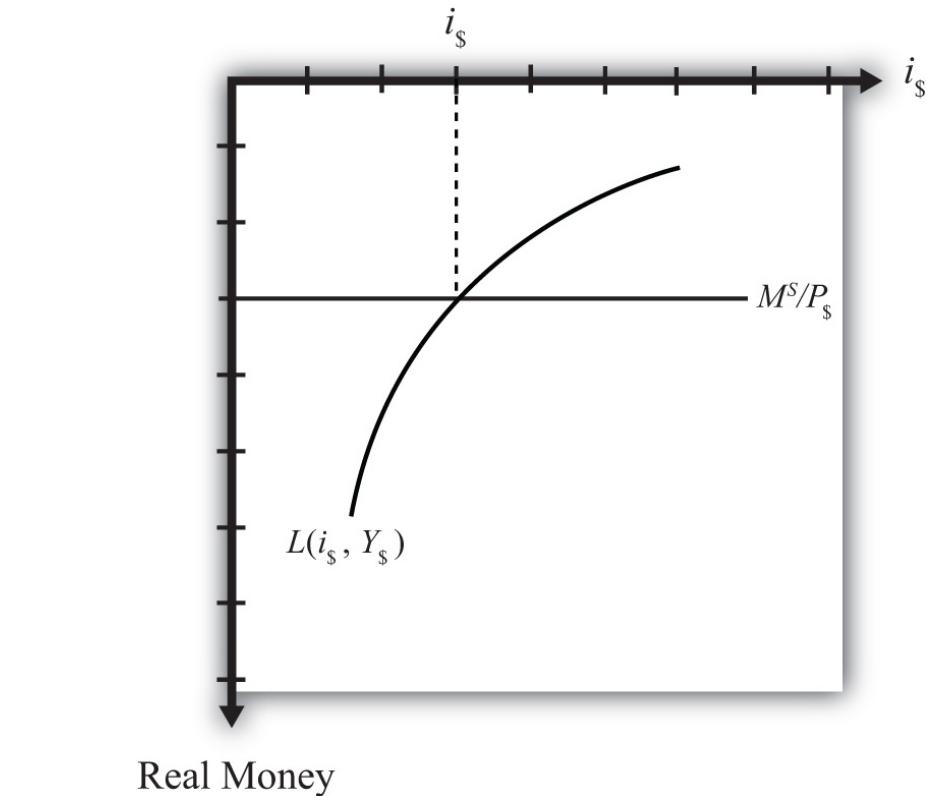
direction. [Figure 18.6 "Rotating the Money Market Diagram"](#) shows the beginning of the rotation pivoted around the origin at zero.

Figure 18.6 Rotating the Money Market Diagram



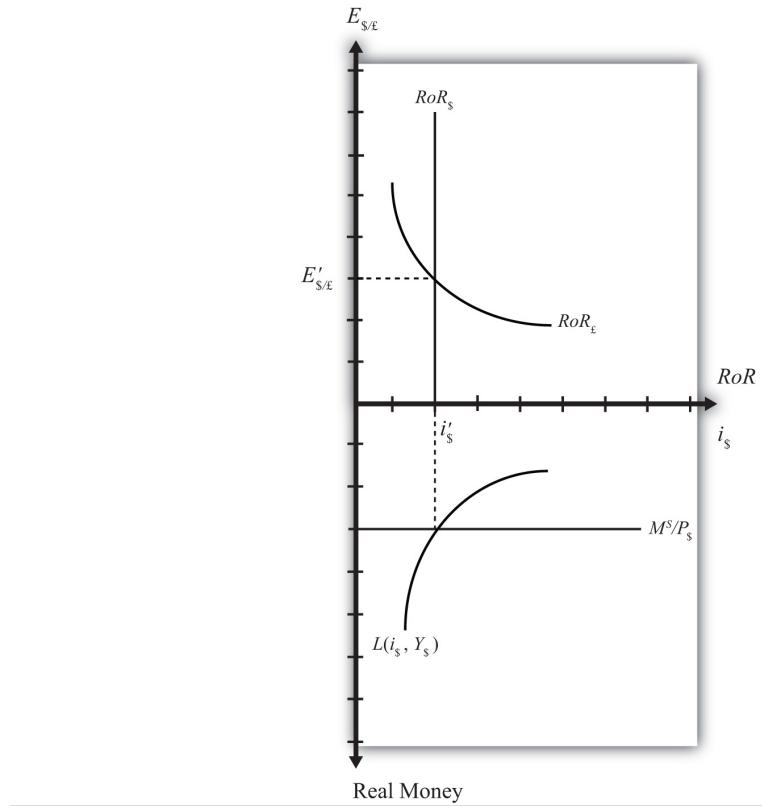
When rotated the full ninety degrees, it will be positioned as shown in [Figure 18.7 "Ninety-Degree Rotation of the Money Market Diagram"](#). The most important thing to remember about this new diagram is that the value of real money supply and demand increases downward away from the origin at zero along the vertical axis. Thus when the money supply “increases,” this will be represented in the diagram as a “downward” shift in the real money supply line. The interest rate, in contrast, increases away from the origin to the right along the horizontal axis when rotated in this position.

Figure 18.7 Ninety-Degree Rotation of the Money Market Diagram



Since the interest rate is identical to the rate of return on dollar assets from a U.S. dollar holder's perspective (i.e., $\text{RoR}_{\$} = i_{\$}$), we can now place the RoR diagram directly on top of the rotated money market diagram as shown in [Figure 18.8 "Money-Forex Diagram"](#). The equilibrium interest rate ($i'_{\$}$), shown along the horizontal axis above the rotated money market diagram, determines the position of the $\text{RoR}_{\$}$ line in the Forex market above. This combined with the $\text{RoR}_{£}$ curve determines the equilibrium exchange rate, $E'_{\$/£}$, in the Forex market. We will call this diagram the "money-Forex diagram" and the combined model the "money-Forex model."

Figure 18.8 Money-Forex Diagram



KEY TAKEAWAY

- Using a two-quadrant diagram with appropriate adjustments, we can represent the equilibrium in the money market and the foreign exchange market simultaneously.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The value of this endogenous variable is used to determine the position of the U.S. rate of return line.
 - b. In the money-Forex diagram, these are the two endogenous variables.
 - c. In the money-Forex diagram, these are the five exogenous variables.

18.13 Comparative Statics in the Combined Money-Forex Model

LEARNING OBJECTIVE

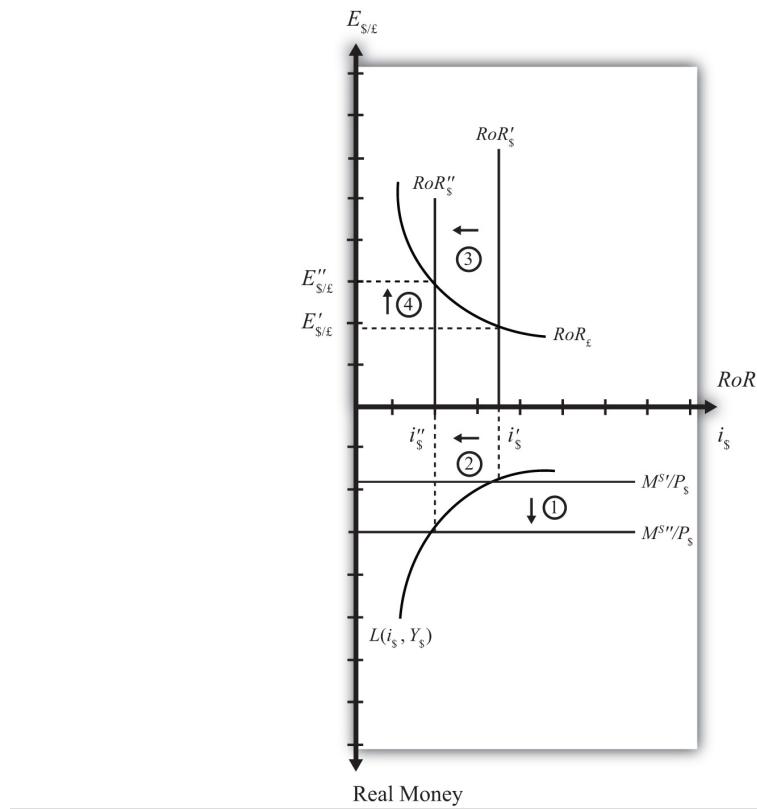
1. Show the effects of an increase in the money supply and an increase in GDP on the interest rate and exchange rate using the two-quadrant money-Forex market diagram.

Comparative statics is any exercise examining how the endogenous variables will be affected when one of the exogenous variables is presumed to change, while holding all other exogenous variables constant. Holding other variables constant at their original values is the “ceteris paribus” assumption. We will do several such exercises here using the combined money-Forex market diagram.

An Increase in the U.S. Money Supply

Suppose the U.S. money supply increases, *ceteris paribus*. The increase in M^S causes an increase in the real money supply ($M^S/P_{\$}$), which causes the real money supply line to shift “down” from $M^{S'}/P_{\$}$ to $M^{S''}/P_{\$}$ (step 1) in the adjacent Money-Forex diagram, [Figure 18.9 "Effects of an Increase in the Money Supply"](#). (Be careful here: down in the diagram means an increase in the real money supply.) This causes a decrease in the equilibrium interest rate from $i_{\$}'$ to $i_{\$}''$ (step 2). The decrease in the U.S. interest rate causes a decrease in the rate of return on dollar assets: $RoR_{\$}$ shifts from $RoR_{\$}'$ to $RoR_{\$}''$ (step 3). Finally, the reduction in the dollar rate of return causes an increase in the exchange rate from $E'_{\$/E}$ to $E''_{\$/E}$ (step 4). This exchange rate change corresponds to an appreciation of the British pound and a depreciation of the U.S. dollar. In summary, an increase in the U.S. money supply, *ceteris paribus*, causes a decrease in U.S. interest rates and a depreciation of the dollar.

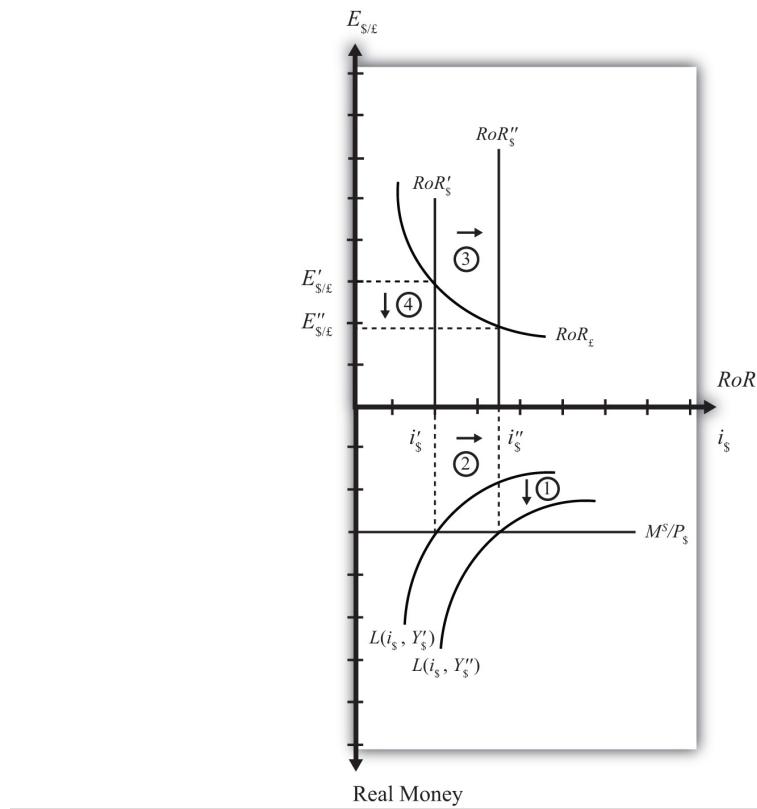
Figure 18.9 Effects of an Increase in the Money Supply



An Increase in U.S. GDP

Suppose there is an increase in U.S. GDP, *ceteris paribus*. This will increase real money demand, causing a “downward” shift in the real money demand curve from $L(i_{\$}, Y_{\$}')$ to $L(i_{\$}, Y_{\$}'')$ (step 1) in the Money-Forex diagram, [Figure 18.10 "Effects of an Increase in GDP"](#). (Remember, real money increases as you move down on the rotated money diagram.) This causes an increase in the U.S. interest rate from $i_{\$}'$ to $i_{\$}''$ (step 2). The increase in the interest means that the rate of return on dollar assets increases from $RoR_{\$}'$ to $RoR_{\$}''$ (step 3). Finally, the increase in the U.S. RoR causes a decrease in the exchange rate from $E'_{\$/£}$ to $E''_{\$/£}$ (step 4). The exchange rate change corresponds to an appreciation of the U.S. dollar and a depreciation of the British pound. In summary, an increase in real U.S. GDP, *ceteris paribus*, causes an increase in U.S. interest rates and appreciation (depreciation) of the U.S. dollar (British pound).

Figure 18.10 Effects of an Increase in GDP



KEY TAKEAWAYS

- In the money-Forex model, an increase in the U.S. money supply, ceteris paribus, causes a decrease in U.S. interest rates and a depreciation of the dollar.
- In the money-Forex model, an increase in real U.S. gross domestic product (GDP), ceteris paribus, causes an increase in U.S. interest rates and appreciation (depreciation) of the U.S. dollar (British pound).

EXERCISE

1. Using the Forex market and money market models, indicate the effect of each change listed in the first row of the table, sequentially, on the variables listed in the first column. For example, “Expansionary U.S. Monetary Policy” will first cause an increase in the “Real U.S. Money Supply.” Therefore, a “+” is placed in the first box of the table. In the next row, answer how “U.S. Interest Rates” will be affected. You do not need to show your work. Note $E_{\$/*}$ represents the dollar/foreign exchange rate. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	Expansionary U.S. Monetary Policy	An Increase in U.S. Price Level	An Increase in U.S. Real GDP
Real U.S. Money Supply	+		
U.S. Interest Rates			
RoR on U.S. Assets			
Foreign Interest Rates			
RoR on Foreign Assets			

	Expansionary U.S. Monetary Policy	An Increase in U.S. Price Level	An Increase in U.S. Real GDP
U.S. Dollar Value			
E/*$			

18.14 Money Supply and Long-Run Prices

LEARNING OBJECTIVE

1. Understand the conditions under which changes in the money supply will have a long-run impact on the price level and hence the inflation rate in a country.

In previous sections we assumed that price levels were given exogenously and were unaffected by changes in other variables. In this section, we will argue that money supply increases tend to have a positive effect on the price level and thus the rate of inflation in an economy. This effect is unlikely to occur quickly, instead arising over several months or years. For this reason, we will say the effect occurs in the *long run*. The magnitude of the price level effect is also greatly influenced by the level of unemployment in the economy. Unemployment affects the degree to which the money increase affects prices and the degree to which it affects output.

The easiest way to see the linkage between money supply and prices is to simplify the story by assuming output cannot change. We tell that in story 1. This assumption allows us to isolate the impact of money on prices alone. In the subsequent adjustment stories, we'll relax the fixed output assumption to show how money increases can also affect the level of output in an economy.

Story 1: Money Supply Increase with Extreme Full Employment

Here we'll consider the effects of a money supply increase assuming what I'll call "extreme full employment." Extreme full employment means that every person who wishes to work within the economy is employed. In addition, each working person is working the maximum number of hours that he or she is willing to work. In terms of capital usage, this too is assumed to be maximally employed. All machinery, equipment, office space, land, and so on that can be employed in production is currently being used. Extreme full employment describes a situation where it is physically impossible to produce any more output with the resources currently available.

Next, let's imagine the central bank increases the money supply by purchasing U.S. government Treasury bills (T-bills) in the open market. Suppose the transaction is made with a commercial bank that decides to sell some of its portfolio of Treasury bills to free reserves to make loans to businesses. The transaction transfers the T-

bill certificate to the central bank in exchange for an accounting notation the central bank makes in the bank's reserve account. Since the transaction increases bank reserves without affecting bank deposits, the bank will now exceed its reserve requirement. Thus these new reserves are available for the bank to lend out.

Let's suppose the value of the T-bills transacted is \$10 million. Suppose the bank decides to lend the \$10 million to Ford Motor Corporation, which is planning to build a new corporate office building. When the loan is made, the bank will create a demand deposit account in Ford's name, which the company can use to pay its building expenses. Only after the creation of the \$10 million demand deposit account is there an actual increase in the money supply.

With money in the bank, Ford will now begin the process of spending to construct the office building. This will involve hiring a construction company. However, Ford will now run into a problem given our assumption of extreme full employment. There are no construction companies available to begin construction on their building. All the construction workers and the construction equipment are already being used at their maximum capacity. There is no leeway.

Nonetheless, Ford has \$10 million sitting in the bank ready to be spent and it wants its building started. So what can it do?

In this situation, the demand for construction services in the economy exceeds the supply. Profit-seeking construction companies that learn that Ford is seeking to begin building as soon as possible, can offer the following deal: "Pay us more than we are earning on our other construction projects and we'll stop working there and come over to build your building." Other construction companies may offer a similar deal. Once the companies, whose construction projects have already started, learn that their construction companies are considering abandoning them for a better offer from Ford, they will likely respond by increasing their payments to their construction crews to prevent them from fleeing to Ford. Companies that cannot afford to raise their payments will be the ones that must cease their construction, and their construction company will flee to Ford. Note that another assumption we must make for this story to work is that there are no enforceable contracts between the construction company and its client. If there were, a company that flees to Ford will find itself being sued for breach of contract. Indeed, this is one of the reasons why contracts are necessary. If all works out perfectly, the least productive construction projects will cease operations since these companies are the ones that are unwilling to raise their wages to keep the construction firm from fleeing.

Once Ford begins construction with its newly hired construction company, several effects are noteworthy. First, Ford's construction company will be working the same amount of time and producing the same amount of output, though for a different client. However, Ford's payments to the construction company are higher now. This means some workers or owners in the construction company are going home with a fatter paycheck. Other construction companies are also receiving higher payments so wages and rents will likely be higher for them as well.

Other companies that have hired the construction firms now face a dilemma, however. Higher payments have to come from somewhere. These firms may respond by increasing the prices of their products for their customers. For example, if this other firm is Coca-Cola, which must now pay higher prices to complete its construction project, it most probably will raise the price of Coke to pay for its higher overall production costs. Hence increases in wages and rents to construction companies will begin to cause increases in market prices of other products, such as Coke, televisions, computers, and so on.

At the same time, workers and owners of the construction companies with higher wages will undoubtedly spend more. Thus they will go out and demand more restaurant meals, cameras, and dance lessons and a whole host of other products. The restaurants, camera makers, and dance companies will experience a slight increase in demand for their products. However, due to the assumption of extreme full employment, they have no ability to increase their supply in response to the increase in demand. Thus these companies will do what the profit-seeking construction companies did before...they will raise their prices.

Thus price increases will begin to ripple through the economy as the extra money enters the circular flow, resulting in demand increases. As prices for final products begin to rise, workers may begin to demand higher wages to keep up with the rising cost of living. These wage increases will in turn lead firms to raise the prices of their outputs, leading to another round of increases in wages and prices. This process is known as the **wage-price spiral**¹⁴.

Nowhere in this process can there ever be more production or output. That's because of our assumption of extreme full employment. We have assumed it is physically impossible to produce any more. For this reason, the only way for the market to reach a new equilibrium with aggregate supply equal to aggregate demand is for prices for most inputs and outputs to rise. In other words, the money supply increase *must* result in an increase in average prices (i.e., the price level) in the economy. Another way of saying this is that money supply increases are inflationary.

14. A process over time in which rising prices cause workers to demand higher wages that in turn leads firms to raise their prices, and so on.

The increase in prices will not occur immediately. It will take time for the construction companies to work out their new payment scheme. It will take more time for them to receive their extra wages and rents and begin spending them. It will take more time, still, for the restaurants and camera makers and others to respond to higher demands. And it will take even more time for workers to respond to the increases in prices and to demand higher wages. The total time may be several years before an economy can get back to equilibrium. For this reason, we think about this money supply effect on the price level as a long-run effect. In other words, we say an increase in the money supply will lead to an increase in the price level in the long run.

Inflation arises whenever there is too much money chasing too few goods. This effect is easy to recognize in this example since output does not change when the money supply increases. So, in this example, there is more money chasing the same quantity of output. Inflation can also arise if there is less output given a fixed amount of money. This is an effect seen in the transition economies of the former Soviet Union. After the breakdown of the political system in the early 1990s, output dropped precipitously, while money in circulation remained much the same. The outcome was a very rapid inflation. In these cases, it was the same amount of money chasing fewer goods.

Story 2: Money Supply Increase with High Unemployment

In this story, we relax the assumption of extreme full employment and assume instead that there is a very high rate of unemployment in the economy. This example will show how money supply increases can affect national output as well as prices.

Suppose there is a money supply increase as in the previous story. When Ford Motor Company goes out looking for a construction company to hire, there is now an important new possibility. Since unemployment is very high, it is likely that most construction companies are not operating at their full capacity. Some companies may have laid off workers in the recent past due to a lack of demand. The construction company that wins the Ford contract will not have to give up other construction projects. Instead, it can simply expand output by hiring unemployed workers and capital. Because there is a ready and waiting source of inputs, even at the original wage and rental rates, there is no need for the construction company to charge Ford more than current prices for its services. Thus there is no pressure to increase wages or the prices of construction services.

It is true, there is more money being paid out in wages by this company, and the new workers will go out and spend that money, leading to an increase in demand

for restaurant services, cameras, dance lessons, and other products. These companies are also likely to respond by hiring more workers and idle equipment to provide more restaurant meals, cameras, and dance lessons. Here too, with a ready and willing source of new inputs from the ranks of the unemployed, these companies will not have an incentive to raise wages, rents, or prices. Instead, they will provide more output of goods and services.

Thus as the increase in money ripples through the economy, it will stimulate demand for a wide variety of products. However, because of high unemployment, the money supply increase need not result in higher prices. Instead, national output increases and the unemployment rate falls.

A comparison of stories 1 and 2 highlights the importance of the unemployment rate in determining the extent to which a money supply increase will be inflationary. In general, we can conclude that an increase in the money supply will raise the domestic price level to a larger degree in the long run, thus lowering the unemployment rate of labor and capital.

Natural Rate of Unemployment

Economists typically say that an economy is at **full employment output**¹⁵ when the unemployment rate is at the **natural rate**¹⁶. The natural rate is defined as the rate that does not cause inflationary pressures in the economy. It is a rate that allows for common transitions that characterize labor markets. For example, some people are currently unemployed because they have recently finished school and are looking for their first job. Some are unemployed because they have quit one job and are in search of another. Some people have decided to move to another city, and are unemployed during the transition. Finally, some people may have lost a job in a company that has closed or downsized and may spend a few weeks or months in search of their next job.

These types of transitions are always occurring in the labor market and are known as *frictional* (or *transitional*) *unemployment*. When employment surveys are conducted each month, they will always identify a group of people unemployed for these reasons. They count as unemployed, since they are all actively seeking work. However, they all will need some time to find a job. As one group of unemployed workers find employment, others will enter the unemployment ranks. Thus there is a constant turnover of people in this group and thus a *natural unemployment rate*. This type of unemployment is also called frictional, or transitional, unemployment. It is distinguished from a second type called structural unemployment. Structural unemployment occurs when there is a change in the structure of production in an economy. For example, if the textile and apparel

15. The level of GDP that would occur when the unemployment rate in the economy is at the natural rate.

16. Defined as the unemployment rate that does not cause inflationary pressures in the economy.

industry closes down and moves abroad, the workers with skills specific to the industry and the capital equipment designed for use in the industry will not be employable in other sectors. These workers and capital may remain unemployed for a longer period of time, or may never find alternative employment.

There is no simple way to measure the natural rate of unemployment. It will likely vary with economic conditions and the fluidity of the labor market. Nonetheless, economists estimate the natural rate of unemployment to be around 5 percent in the United States today.

When economists talk about the inflationary effect of money supply increases, they typically refer to the natural rate of unemployment. A money supply increase will likely be inflationary when the unemployment rate is below the natural rate. In contrast, inflationary effects of money supply increases are reduced if the economy has unemployment above the natural rate. Here's how the story would work.

Story 3: Money Supply Increase above and below the Natural Unemployment Rate

Suppose there is a money supply increase as in the previous story, but now let's assume the economy is operating above full employment, meaning that unemployment is below its natural rate.

As the money supply increase ripples through the economy causing excess demand, as described above, businesses have some leeway to expand output. Since unemployment is not zero, they can look to hire unemployed workers and expand output. However, as frictional unemployment decreases, the labor market will pick up speed. Graduating students looking for their first job will find one quickly. Workers moving to another job will also find one quickly. In an effort to get the best workers, firms may begin to raise their wage offers. Workers in transition may quickly find themselves entertaining several job offers, rather than just one. These workers will begin to demand higher wages. Ultimately, higher wages and rents will result in higher output prices, which in turn will inspire demands for higher wages. Thus despite the existence of some unemployment, the money supply increase may increase output slightly but it is also likely to be inflationary.

In contrast, suppose the economy were operating with unemployment above the natural rate. In this case, the increase in demand caused by a money supply increase is likely to have a more significant effect upon output. As firms try to expand output, they will face a much larger pool of potential employees. Competition by several workers for one new job will put power back in the hands of

the company, allowing it to hire the best quality worker without having to raise its wage offer to do so. Thus, in general, output will increase more and prices will increase less, if at all. Thus the money supply increase is less likely to be inflationary in the long run when the economy is operating above the natural rate of unemployment.

KEY TAKEAWAYS

- Inflation arises whenever there is too much money chasing too few goods.
- A money supply increase will lead to increases in aggregate demand for goods and services.
- A money supply increase will tend to raise the price level in the long run.
- A money supply increase may also increase national output.
- A money supply increase will raise the price level more and national output less the lower the unemployment rate of labor and capital is.
- A money supply increase will raise national output more and the price level less the higher the unemployment rate of labor and capital is.
- The natural rate of unemployment is the rate that accounts for frictional unemployment. It is also defined as the rate at which there are no aggregate inflationary pressures.
- If a money supply increase drives an economy below the natural rate of unemployment, price level increases will tend to be large while output increases will tend to be small.
- If a money supply increase occurs while an economy is above the natural rate of unemployment, price level increases will tend to be small while output increases will tend to be large.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term coined in this text for the situation when everybody who wishes to work is employed.
 - b. The term used to describe how increases in output prices lead to increases in wages, which further cause output prices to rise ad infinitum.
 - c. The term for the unemployment rate at which there is no inflationary or deflationary pressure on average prices.
 - d. The term for the level of GDP in an economy when the unemployment rate is at its natural level.
 - e. The term used to describe the type of unemployment that arises because of the typical adjustments of workers into, out of, and between jobs in an economy.
 - f. The likely larger long-run effect of a money supply increase when an economy has unemployment below the natural rate.
 - g. The likely larger long-run effect of a money supply increase when an economy has unemployment above the natural rate.

Chapter 19

National Output Determination

In most introductory macroeconomics courses, the basic Keynesian model is presented as a way of showing how government spending and taxation policies can influence the size of a country's growth national product (GNP). This chapter revisits the basic Keynesian model but adds an international angle by including impacts on domestic demand for goods and services caused by changes in the exchange rate. With this relationship in place, the chapter concludes with several comparative statics exercises showing how changes in key variables may influence the level of GNP.

19.1 Overview of National Output Determination

LEARNING OBJECTIVE

- Understand the structure and results of the basic Keynesian model of national output determination.

This chapter describes how the supply and demand for the national output of goods and services combine to determine the equilibrium level of national output for an economy. The model is called the goods and services market model, or just the G&S market model.

In this model, we use gross national product (GNP) as the measure of national output rather than gross domestic product (GDP). This adjustment is made because we wish to define the trade balance ($EX - IM$) as the current account (defined as the difference between exports and imports of goods, services incomes payments/receipts, and unilateral transfers). This adjustment is discussed in more detail in [Section 19.6 "Export and Import Demand"](#).

The diagram used to display this model is commonly known as the Keynesian cross. The model assumes, for simplicity, that the amount of national output produced by an economy is determined by the total amount demanded. Thus if, for some reason, the demand for GNP were to rise, then the amount of GNP supplied would rise up to satisfy it. If demand for the GNP falls—for whatever reason—then supply of GNP would also fall. Consequently, it is useful to think of this model as “demand driven.”

The model is developed by identifying the key determinants of GNP demand. The starting point is the national income identity, which states that

$$GNP = C + I + G + EX - IM,$$

that is, the gross national product is the sum of consumption expenditures (C), investment expenditures (I), government spending (G), and exports (EX) minus imports (IM).

Note that the identity uses GNP rather than GDP if we define EX and IM to include income payments, income receipts, and unilateral transfers as well as goods and services trade.

We rewrite this relationship as

$$AD = C^D + I^D + G^D + EX^D - IM^D,$$

where AD refers to aggregate demand for the GNP and the right-side variables are now read as consumption demand, investment demand, and so on. The model further assumes that consumption demand is positively related to changes in **disposable income**¹ (Y_d). Furthermore, since disposable income is in turn negatively related to taxes and positively related to transfer payments, these additional variables can also affect aggregate demand.

The model also assumes that demand on the current account ($CA^D = EX^D - IM^D$) is negatively related to changes in the domestic real currency value (i.e., the real exchange rate) and changes in disposable income. Furthermore, since the domestic real currency value is negatively related to the domestic price level (inflation) and positively related to the foreign price level, these variables will also affect current account demand.

Using the G&S market model, several important relationships between key economic variables are shown:

- When government demand (G) or investment demand (I) for G&S rises (falls), equilibrium GNP rises (falls).
- When disposable income rises (falls) due to a decrease (increase) in taxes or an increase (decrease) in transfer payments, equilibrium GNP increases (decreases).
- When the real exchange rate depreciates (appreciates), either due to a depreciation of the nominal exchange rate, an increase in the domestic price level, or a decrease in the foreign price level, equilibrium GNP rises (falls).

Connections

The G&S market model connects with the money market because the value of GNP determined in the G&S model affects money demand. If equilibrium GNP rises in the G&S model, then money demand will rise, causing an increase in the interest rate.

1. All the income households have at their disposal to spend. It is defined as national income (GNP), minus taxes taken away by the government, plus transfer payments that the government pays out to people.

The G&S model also connects with the foreign exchange (Forex) market. The equilibrium exchange rate determined in the Forex affects the real exchange rate that in turn influences demand on the current account.

A thorough discussion of these interrelationships is given in [Chapter 20 "The AA-DD Model"](#).

Omissions

There is one important relationship omitted in this version of the G&S model, and that is the relationship between interest rates and investment. In most standard depictions of the Keynesian G&S model, it is assumed that increases (decreases) in interest rates will reduce (increase) demand for investment. In this version of the model, to keep things simple, investment is assumed to be exogenous (determined in an external process) and unrelated to the level of interest rates.

Some approaches further posit that interest rates affect consumption demand as well. This occurs because household borrowing, to buy new cars or other consumer items, will tend to rise as interest rates fall. However, this relationship is also *not* included in this model.

KEY TAKEAWAYS

- The Keynesian, or G&S, model of output determination is a demand-driven model in that the amount of national output produced by an economy is determined by the total amount demanded.
- One important relationship omitted in this version of the G&S model is the lack of a relationship between interest rates and investment.
- The main results from the G&S model are the following:
 - When government demand (G) or investment demand (I) for G&S rises (falls), equilibrium GNP rises (falls).
 - When disposable income rises (falls) due to a decrease (increase) in taxes or an increase (decrease) in transfer payments, equilibrium GNP increases (decreases).
 - When the real exchange rate depreciates (appreciates), either due to a depreciation of the nominal exchange rate, an increase in the domestic price level, or a decrease in the foreign price level, equilibrium GNP rises (falls).

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. In the Keynesian, or G&S, model, this is the primary determinant of aggregate supply.
 - b. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP if government spending decreases in the G&S model.
 - c. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP if investment spending increases in the G&S model.
 - d. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP if tax revenue decreases in the G&S model.
 - e. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP if transfer payments increase in the G&S model.
 - f. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP if the domestic currency depreciates in the G&S model.
 - g. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP if the domestic price level decreases in the G&S model.
 - h. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP if the foreign price level decreases in the G&S model.

19.2 Aggregate Demand for Goods and Services

LEARNING OBJECTIVE

1. Learn that aggregate demand is the summation of the separate demands for each variable in the national income identity.

The Keynesian model of aggregate demand for goods and services is developed by identifying key determinants of demand for the national output. When we talk about *aggregate* demand (AD), it means demand by households, businesses, and the government for anything and everything produced within the economy. The starting point is the national income identity, which states that

$$GNP = C + I + G + EX - IM,$$

that is, the gross national product is the sum of consumption expenditures, investment expenditures, government spending, and exports minus imports of goods and services.

We rewrite this relationship as

$$AD = C^D + I^D + G^D + EX^D - IM^D,$$

where the left side, AD, refers to aggregate demand for the GNP and the right-side variables are read as consumption demand, investment demand, and so on. Determinants of the right-side variables will be considered in turn.

It is important to remember that demand is merely what households, businesses, and the government “would like” to purchase given the conditions that exist in the economy. Sometimes demand will be realized, as when the economy is in equilibrium, but sometimes demand will not be satisfied. On the other hand, the variable Y, for real GNP, represents the aggregate supply of G&S. This will correspond to the actual GNP whether in equilibrium or not.

Next, we’ll present the determinants of each demand term: consumption, investment, government, and export and import demand.

KEY TAKEAWAY

- In the G&S model, aggregate demand for the GNP is the sum of consumption demand, investment demand, government demand, and current account demand.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. In the G&S model, the variable Y stands for this.
 - b. In the G&S model, the variable AD stands for this.
 - c. In the G&S model, the variable I^D stands for this.
 - d. In the G&S model, the variable EX^D stands for this.
 - e. In the G&S model, the variable CA^D stands for this.

19.3 Consumption Demand

LEARNING OBJECTIVE

1. Learn the determinants of consumption demand and the effects of changes in these variables.

Consumption demand represents the demand for goods and services by individuals and households in the economy. This is the major category in the national income accounts for most countries, typically comprising from 50 percent to 70 percent of the gross national product (GNP) for most countries.

In this model, the main determinant of consumption demand is disposable income. Disposable income is all the income households have at their disposal to spend. It is defined as national income (GNP) minus taxes taken away by the government, plus **transfer payments**² that the government pays out to people. More formally, this is written as

$$Y_d = Y - T + TR,$$

where Y_d refers to disposable income, Y is real GNP, T is taxes, and TR represents transfer payments.

In this relationship, disposable income is defined in the same way as in the circular flow diagram presented in [Chapter 13 "National Income and the Balance of Payments Accounts"](#), [Section 13.7 "The Twin-Deficit Identity"](#). Recall that taxes withdrawn from GNP are assumed to be all taxes collected by the government from all sources. Thus income taxes, social insurance taxes, profit taxes, sales taxes, and property taxes are all assumed to be included in taxes (T). Also, transfer payments refer to all payments made by the government that do not result in the provision of a good or service. All social insurance payments, welfare payments, and unemployment compensation, among other things, are included in transfers (TR).

2. Government expenditures made to individuals or businesses without any good or service received in exchange. It includes social insurance payments, welfare payments and unemployment compensation.

In the G&S model, demand for consumption G&S is assumed to be positively related to disposable income. This means that when disposable income rises, demand for consumption G&S will also rise, and vice versa. This makes sense since households who have more money to spend will quite likely wish to buy more G&S.

We can write consumption demand in a functional form as follows:

$$C^D(Y_d) = C^D(Y - T + TR).$$

This expression says that consumption demand is a function C^D that depends positively (+) on disposable income (Y_d). The second term simply substitutes the variables that define disposable income in place of Y_d . It is a more complete way of writing the function. Note well that C^D here denotes a function, not a variable. The expression is the same as if we had written $f(x)$, but instead we substitute a C^D for the f and Y_d for the x .

It is always important to keep track of which variables are exogenous and which are endogenous. In this model, real GNP (Y) is the key endogenous variable since it will be determined in the equilibrium. Taxes (T) and transfer payments (TR) are exogenous variables, determined outside the model. Since consumption demand C^D is dependent on the value of Y , which is endogenous, C^D is also endogenous. By the same logic, Y_d is endogenous as well.

Linear Consumption Function

It is common in most introductory textbooks to present the consumption function in linear form. For our purposes here, this is not absolutely necessary, but doing so will allow us to present a few important points.

In linear form, the consumption function is written as

$$C^D = C_0 + mpcY_d = C_0 + mpc(Y - T + TR).$$

Here C_0 represents autonomous consumption and mpc refers to the *marginal propensity to consume*.

Autonomous consumption (C_0) is the amount of consumption that would be demanded even if income were zero. (Autonomous simply means “independent” of income.) Graphically, it corresponds to the y -intercept of the linear function. Autonomous consumption will be positive since households will spend some money (drawing on savings if necessary) to purchase consumption goods (like food) even if income were zero.

The **marginal propensity to consume (mpc)**³ represents the additional (or marginal) demand for G&S given an additional dollar of disposable income. Graphically, it corresponds to the slope of the consumption function. This variable must be in the range of zero to one and is most likely to be between 0.5 and 0.8 for most economies. If mpc were equal to one, then households would spend every additional dollar of income. However, because most households put some of their income into savings (i.e., into the bank, or pensions), not every extra dollar of income will lead to a dollar increase in consumption demand. That fraction of the dollar not used for consumption but put into savings is called the **marginal propensity to save (mps)**⁴. Since each additional dollar must be spent or saved, the following relationship must hold:

$$mpc + mps = 1,$$

that is, the sum of the marginal propensity to consume and the marginal propensity to save must equal 1.

KEY TAKEAWAYS

- In the G&S model, consumption demand is determined by disposable income.
- A linear consumption function includes the marginal propensity to consume and an autonomous consumption component, besides disposable income.
- Disposable income is defined as national income (GNP) minus taxes plus transfer payments.
- An increase (decrease) in disposable income will cause an increase (decrease) in consumption demand.
- An increase (decrease) in the marginal propensity to consume will cause an increase (decrease) in consumption demand.

3. The additional consumption demanded with an additional dollar of disposable income.

4. The additional saving that occurs with an additional dollar of disposable income.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term that represents the additional amount of consumption demand caused by an additional dollar of disposable income.
 - b. The term that represents the additional amount of saving caused by an additional dollar of disposable income.
 - c. The term for the amount of consumption demand that would arise even if disposable income were zero.
 - d. Of *positive* or *negative*, the relationship between changes in disposable income and changes in consumption demand.
 - e. Of *positive* or *negative*, the relationship between changes in tax revenues and changes in consumption demand.
 - f. Of *positive* or *negative*, the relationship between changes in real GNP and changes in consumption demand.
 - g. A household purchase of a refrigerator would represent demand recorded in this component of aggregate demand in the G&S model.

19.4 Investment Demand

LEARNING OBJECTIVE

1. Learn the determinants of investment demand and the effects of changes in these variables.

Investment demand refers to the demand by businesses for physical capital goods and services used to maintain or expand its operations. Think of it as the office and factory space, machinery, computers, desks, and so on that are used to operate a business. It is important to remember that investment demand here does not refer to financial investment. Financial investment is a form of saving, typically by households that wish to maintain or increase their wealth by deferring consumption till a later time.

In this model, investment demand will be assumed to be exogenous. This means that its value is determined outside of the model and is not dependent on any variable within the model. This assumption is made primarily to simplify the analysis and to allow the focus to be on exchange rate changes later. The simple equation for investment demand can be written as

$$I^D = I_0,$$

where the “0,” or naught, subscript on the right side indicates that the variable is exogenous or autonomous. In words, the equation says that investment demand is given exogenously as I_0 .

Admittedly, this is not a realistic assumption. In many other macro models, investment demand is assumed to depend on two other aggregate variables: GNP and interest rates. GNP may affect investment demand since the total demand for business expansion is more likely the higher the total size of the economy. The growth rate of GNP may also be an associated determinant since the faster GNP is growing, the more likely companies will predict better business in the future, inspiring more investment.

Interest rates can affect investment demand because many businesses must borrow money to finance expansions. The interest rate is the cost of borrowing money;

thus, the higher the interest rates are, the lower the investment demand should be, and vice versa.

If we included the GNP and interest rate effects into the model, the solution to the extended model later would prove to be much more difficult. Thus we simplify things by assuming that investment is exogenous. Since many students have learned about the GNP and interest rate effect in previous courses, you need to remember that these effects are *not* a part of this model.

KEY TAKEAWAYS

- In the G&S model, investment demand is assumed to be exogenous, meaning not dependent on any other variable within the model such as GNP or interest rates.
- The omission of an effect by GNP and interest rates on investment demand is made to simplify the model.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Term for a type of investment by households that wish to maintain or increase their wealth by deferring consumption till a later time.
 - b. Investment demand refers to this type of goods and services.
 - c. Of *exogenous* or *endogenous*, this describes investment demand in the G&S model in the text.
 - d. The name of two variables that are likely to influence investment demand in reality but are excluded from the G&S model as a simplification.
 - e. A business purchase of a company delivery van would represent demand recorded in this component of aggregate demand in the G&S model.

19.5 Government Demand

LEARNING OBJECTIVE

1. Learn the determinants of government demand and the effects of changes in these variables.

Government demand refers to the government's demand for goods and services produced in the economy. In some cases this demand is for G&S produced by private businesses, as when the government purchases a naval aircraft. Other government demand is actually produced by the government itself, as what occurs with teachers providing educational services in the public schools. All levels of government demand—federal, state, and local—are included in this demand term. Excluded are transfer payments such as social insurance, welfare assistance, and unemployment compensation.

In this model, government demand is treated the same way as investment demand: it is assumed to be exogenous. This means that its value is determined outside of the model and is not dependent on any variable within the model. A simple equation for government demand can be written as

$$G^D = G_0,$$

where the “0,” or naught, subscript on the right side indicates that the variable is exogenous or autonomous. In words, the equation says that government demand is given exogenously as G_0 .

This is a more common assumption in many other macro models, even though one could argue dependencies of government demand on GNP and interest rates. However, these linkages are not likely to be as strong as with investment, thus assuming exogeneity here is a more realistic assumption than with investment.

KEY TAKEAWAY

- In the G&S model, government demand is assumed to be exogenous, meaning not dependent on any other variable within the model such as GNP or interest rates.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. These three levels of government demand are included in G^D in the G&S model.
 - b. This type of government expenditure is not included in G^D in the G&S model.
 - c. Of *exogenous* or *endogenous*, this describes government demand in the G&S model in the text.
 - d. An expenditure by a state school system on teachers’ salaries would represent demand recorded in this component of aggregate demand in the G&S model.

19.6 Export and Import Demand

LEARNING OBJECTIVE

1. Learn the determinants of export and import demand and the effects of changes in these variables.

Export demand refers to the demand by foreign countries for G&S produced domestically. Ultimately, these goods are exported to foreign residents. Import demand refers to demand by domestic residents for foreign-produced G&S. Imported G&S are not a part of domestic GNP. Recall from [Chapter 13 "National Income and the Balance of Payments Accounts"](#), [Section 13.3 "U.S. National Income Statistics \(2007–2008\)"](#) that imports are subtracted from the national income identity because they are included as a part of consumption, investment, and government expenditures as well as in exports. Likewise in this model, consumption, investment, government, and export demand are assumed to include demand for imported goods. Thus imports must be subtracted to assure that only domestically produced G&S are included.

We will define current account demand as $CA^D = EX^D - IM^D$. The key determinants of current account demand in this model are assumed to be the domestic real currency value and disposable income.

First, let's define the real currency value, then show how it relates to the demand for exports and imports. The real British pound value in terms of U.S. dollars (also called the **real exchange rate**⁵ between dollars and pounds), $RE_{\$/\text{£}}$, is a measure of the cost of a market basket of goods abroad relative to the cost of a similar basket domestically. It captures differences in prices, converted at the spot exchange rate, between the domestic country and the rest of the world. It is defined as

$$RE_{\$/\text{£}} = \frac{E_{\$/\text{£}} CB_{\text{£}}}{CB_{\$}},$$

5. A measure of the cost of a market basket of goods abroad relative to the cost of a similar basket domestically. It captures differences in prices, converted at the spot exchange rate, between the domestic country and the rest of the world.

where $E_{\$/\text{£}}$ is the spot exchange rate, $CB_{\text{£}}$ is the cost of a market basket of goods in Britain, and $CB_{\$}$ is the cost of a comparable basket of goods in the United States. The top expression, $E_{\$/\text{£}} CB_{\text{£}}$, represents the cost of a British market basket of goods converted to U.S. dollars. Thus if $RE_{\$/\text{£}} > 1$, then a British basket of goods costs more than a comparable U.S. basket of goods. If $RE_{\$/\text{£}} < 1$, then a U.S. basket of goods costs

more than a British basket. Also note that $RE_{\$/\text{£}}$ is a unitless number. If $RE_{\$/\text{£}} = 2$, for example, it means that British goods cost twice as much as U.S. goods, on average, at the current spot exchange rate.

Note that we could also have defined the reciprocal real exchange rate, $RE_{\text{£}/\$}$. This real exchange rate is the real value of the pound in terms of U.S. dollars. Since the real exchange rate can be defined in two separate ways between any two currencies, it can be confusing to say things like “the real exchange rate rises” since the listener may not know which real exchange rate the speaker has in mind. Thus it is always preferable to say the real dollar value rises, or the real pound value falls, since this eliminates any potential confusion. In this text, I will always adhere to the convention of writing the spot exchange rate and the real exchange rate with the U.S. dollar in the numerator. Thus references to the real exchange rate in this text will always refer to that form.

Since the cost of a market basket of goods is used to create a country’s price index, changes in CB will move together with changes in the country’s price level P . For this reason, it is common to rewrite the real exchange rate using price levels rather than costs of market baskets and to continue to interpret the expression in the same way. For more information related to this, see [Chapter 17 "Purchasing Power Parity"](#), [Section 17.2 "The Consumer Price Index \(CPI\) and PPP"](#). We will follow that convention here and rewrite $RE_{\$/\text{£}}$ as

$$RE_{\$/\text{£}} = \frac{E_{\$/\text{£}} P_{\text{£}}}{P_{\$}},$$

where $P_{\text{£}}$ is the British price index and $P_{\$}$ is the U.S. price index. From this point forward, we’ll mean this expression whenever we discuss the real exchange rate.

Next, we’ll discuss the connection to current account demand. To understand the relationship it is best to consider a change in the real exchange rate. Suppose $RE_{\$/\text{£}}$ rises. This means that the real value of the pound rises and, simultaneously, the real U.S. dollar value falls. This also means that goods and services are becoming relatively more expensive, on average, in Britain compared to the United States.

Relatively cheaper G&S in the United States will tend to encourage U.S. exports since the British would prefer to buy some cheaper products in the United States. Similarly, relatively more expensive British G&S will tend to discourage U.S. imports from Britain. Since U.S. exports will rise and imports will fall with an increase in the real U.S. dollar value, current account demand, $CA^D = EX^D - IM^D$, will rise. Similarly, if the real U.S. dollar value falls, U.S. exports will fall and imports

rise, causing a decrease in CA^D . Hence, there is a positive relationship between this real exchange rate (i.e., the real value of the pound) and U.S. current account demand.

Disposable income can also affect the current account demand. In this case, the effect is through imports. An increase in disposable income means that households have more money to spend. Some fraction of this will be consumed, the marginal propensity to consume, and some fraction of that consumption will be for imported goods. Thus an increase in disposable income should result in an increase in imports and a subsequent reduction in current account demand. Thus changes in disposable income are negatively related to current account demand.

We can write current account demand in functional form as follows:

$$CA^D(RE_{\$/\text{£}}^+, Y_d^-).$$

The expression indicates that current account demand is a function of $RE_{\$/\text{£}}$ and Y_d . The “+” sign above $RE_{\$/\text{£}}$ indicates the positive relationship between the real exchange rate (as defined) and current account demand. The “−” sign above the disposable income term indicates a negative relationship with current account demand.

KEY TAKEAWAYS

- The key determinants of current account demand in the G&S model are assumed to be the domestic real currency value and disposable income.
- The real exchange rate captures differences in prices, converted at the spot exchange rate, between the domestic country and the rest of the world.
- In the G&S model, there is a positive relationship between the real exchange rate (as defined) and current account demand and a negative relationship between disposable income and current account demand.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *positive or negative*, the relationship between changes in the domestic price level and the real value of the domestic currency.
 - b. Of *positive or negative*, the relationship between changes in the foreign price level and the real value of the domestic currency.
 - c. Of *positive or negative*, the relationship between changes in the nominal value of the domestic currency and the real value of the domestic currency.
 - d. Of *increase, decrease, or stay the same*, the effect of a real appreciation of the domestic currency on current account demand in the G&S model.
 - e. Of *increase, decrease, or stay the same*, the effect of a depreciation of the domestic currency on current account demand in the G&S model.
 - f. Of *increase, decrease, or stay the same*, the effect of an increase in the domestic price level on current account demand in the G&S model.
 - g. Of *increase, decrease, or stay the same*, the effect of an increase in the foreign price level on current account demand in the G&S model.
 - h. Of *increase, decrease, or stay the same*, the effect of a decrease in real GNP on current account demand in the G&S model.
 - i. An expenditure by a domestic business for a microscope sold by a Swiss firm would represent demand recorded in this component of aggregate demand in the G&S model.
 - j. An expenditure by a foreign business for a microscope sold by a U.S. firm would represent demand recorded in this component of aggregate demand in the G&S model.

19.7 The Aggregate Demand Function

LEARNING OBJECTIVE

1. Combine the individual demand functions into an aggregate demand (AD) function.

Notice that the right side indicates that if disposable income were to rise, consumption demand would rise but current account demand, which is negatively related to disposable income, would fall. This would seem to make ambiguous the effect of a disposable income change on aggregate demand. However, by thinking carefully about the circular flow definitions, we can recall that consumption expenditures consist of the sum of expenditures on domestically produced goods and imported goods. This was the reason imports are subtracted away in the national income identity. This also means that the marginal propensity to spend on imported goods must be lower than the total marginal propensity to consume, again since imports are a fraction of total consumption spending. This implies that the negative effect on imports from a \$1 increase in disposable income must be less than the positive impact on consumption demand.

We indicate the net positive effect on aggregate demand of changes in disposable income with the “+” sign above Y_d on the left-hand side. The positive impact of changes in the real exchange rate, investment demand, and government demand is obvious and is also shown.

We can write the aggregate demand function in several different ways. To be more explicit, we can include all the fundamental variables affecting aggregate demand by writing out the disposable income and real exchange rate terms as follows:

$$AD(Y - T^+ + TR, \frac{E_{\$/\text{£}} P_{\text{£}}}{P_{\$}}, I_0^+, G_0^+) = C^D(Y - T^+ + TR) + I_0 + G_0 + CA^D$$

Writing the expression in this way allows us to indicate that the spot exchange rate, the price levels domestically and abroad, and domestic taxes and transfer payments also affect aggregate demand. For example, increases in autonomous transfer payments will raise aggregate demand since it raises disposable income, which in turn raises demand. Increases in taxes, however, will lower disposable income, which in turn will lower aggregate demand. Similarly, an increase in the spot

exchange rate (as defined) or the foreign price level will raise aggregate demand, since both changes will increase the real exchange rate. However, an increase in the domestic price level will reduce the real exchange rate (because it is in the denominator) and thus reduce aggregate demand.

KEY TAKEAWAY

- Aggregate demand is positively related to changes in disposable income, the real exchange rate (as defined), and investment and government demands.

EXERCISE

- 1. Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - Of *increase, decrease, or stay the same*, the effect of a real appreciation of the domestic currency on aggregate demand in the G&S model.
 - Of *increase, decrease, or stay the same*, the effect of an increase in investment demand on aggregate demand in the G&S model.
 - Of *increase, decrease, or stay the same*, the effect of an increase in disposable income on aggregate demand in the G&S model.
 - Of *increase, decrease, or stay the same*, the effect of an increase in income taxes on aggregate demand in the G&S model.
 - Of *increase, decrease, or stay the same*, the effect of an increase in government demand on aggregate demand in the G&S model.
 - Of *increase, decrease, or stay the same*, the effect of an increase in the real currency value on aggregate demand in the G&S model.
 - Of *increase, decrease, or stay the same*, the effect of an increase in the domestic price level on aggregate demand in the G&S model.

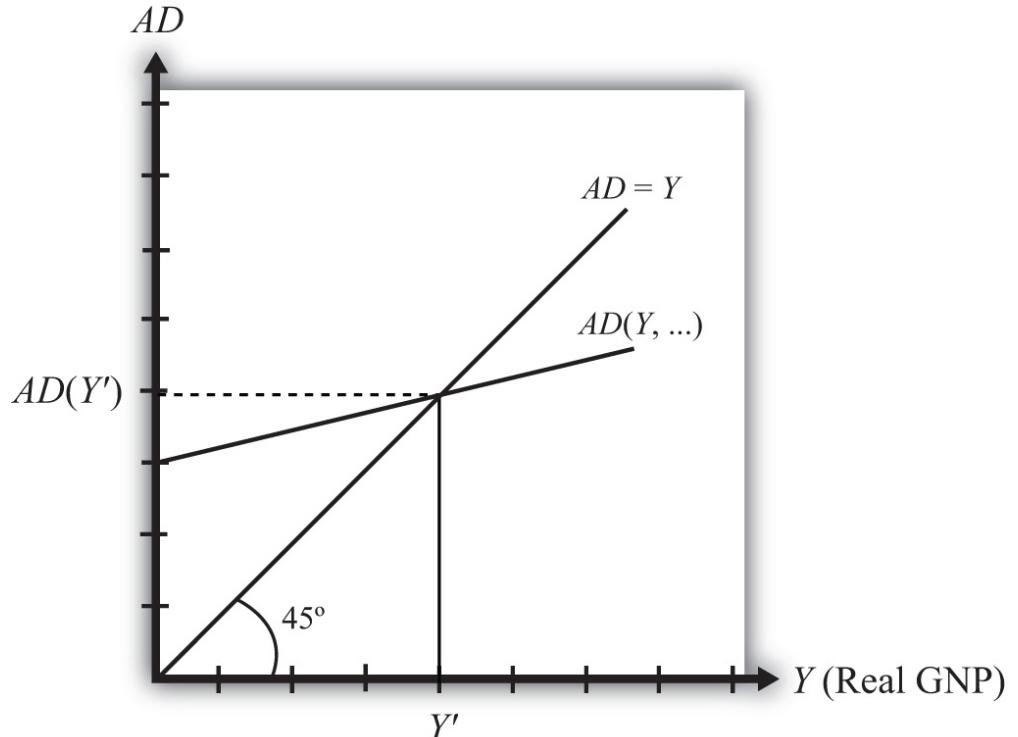
19.8 The Keynesian Cross Diagram

LEARNING OBJECTIVE

1. Learn how to use the Keynesian cross diagram to describe equilibrium in the G&S market.

The **Keynesian cross diagram**⁶ depicts the equilibrium level of national income in the G&S market model. We begin with a plot of the aggregate demand function with respect to real GNP (Y) in [Figure 19.1 "Aggregate Demand Function"](#). Real GNP (Y) is plotted along the horizontal axis, and aggregate demand is measured along the vertical axis. The aggregate demand function is shown as the upward sloping line labeled $AD(Y, \dots)$. The (...) is meant to indicate that AD is a function of many other variables not listed. There are several important assumptions about the form of the AD function that are needed to assure an equilibrium. We discuss each assumption in turn.

Figure 19.1 Aggregate Demand Function



6. Depicts the equilibrium level of national income in the G&S market model.

First, the AD function is positively sloped with respect to changes in Y , ceteris paribus. Recall that ceteris paribus means that all other variables affecting aggregate demand are assumed to remain constant as GNP changes. The positive slope arises from the rationale given previously that an increase in disposable income should naturally lead to an increase in consumption demand and a smaller decrease in CA demand, resulting in a net increase in aggregate demand. Next, if GNP rises, ceteris paribus, it means that taxes and transfer payments remain fixed and disposable income must increase. Thus an increase in GNP leads to an increase in AD.

Second, the AD function has a positive vertical intercept term. In other words, the AD function crosses the vertical axis at a level greater than zero. For reasons that are not too important, this feature is critical for generating the equilibrium later. The reason it arises is because autonomous consumption, investment, and government demand are all assumed to be independent of income and positive in value. These assumptions guarantee a positive vertical intercept.

Third, the AD function has a slope that is less than one. This assumption means that for every \$1 increase in GNP (Y), there is a less than \$1 increase in aggregate demand. This arises because the marginal propensity to consume domestic GNP is less than one for two reasons. First, some of the additional income will be spent on imported goods, and second, some of the additional income will be saved. Thus the AD function will have a slope less than one.

Also plotted in the diagram is a line labeled $AD = Y$. This line is also sometimes called the forty-five-degree line since it sits at a forty-five-degree angle to the horizontal axis. This line represents all the points on the diagram where AD equals GNP. Since GNP can be thought of as aggregate supply, the forty-five-degree line contains all the points where AD equals aggregate supply.

Because of the assumptions about the shape and position of the AD function, AD will cross the forty-five-degree line, only once, from above. The intersection determines the equilibrium value of GNP, labeled Y' in the diagram.

KEY TAKEAWAYS

- The Keynesian cross diagram plots the aggregate demand function versus GNP together with a forty-five-degree line representing the set of points where $AD = GNP$. The intersection of these two lines represents equilibrium GNP in the economy.
- An equilibrium exists if the AD function crosses the forty-five-degree line from above. This occurs if three conditions hold:
 1. The AD function has a positive slope. (It does.)
 2. The AD function has a slope less than one. (It does.)
 3. The AD function intersects the vertical axis in the positive range. (It does.)

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *positive, negative, or zero*, the slope of an aggregate demand function with respect to changes in real GNP.
 - b. Of *positive, negative, or zero*, the value of the vertical intercept of an aggregate demand function.
 - c. Of *greater than one, less than one, or equal to one*, the value of the slope of an aggregate demand function with respect to changes in real GNP.
 - d. The equality that is satisfied on the forty-five-degree line in a Keynesian cross diagram.
 - e. The value of this variable is determined at the intersection of the aggregate demand function and the forty-five-degree line in a Keynesian cross diagram.

19.9 Goods and Services Market Equilibrium Stories

LEARNING OBJECTIVE

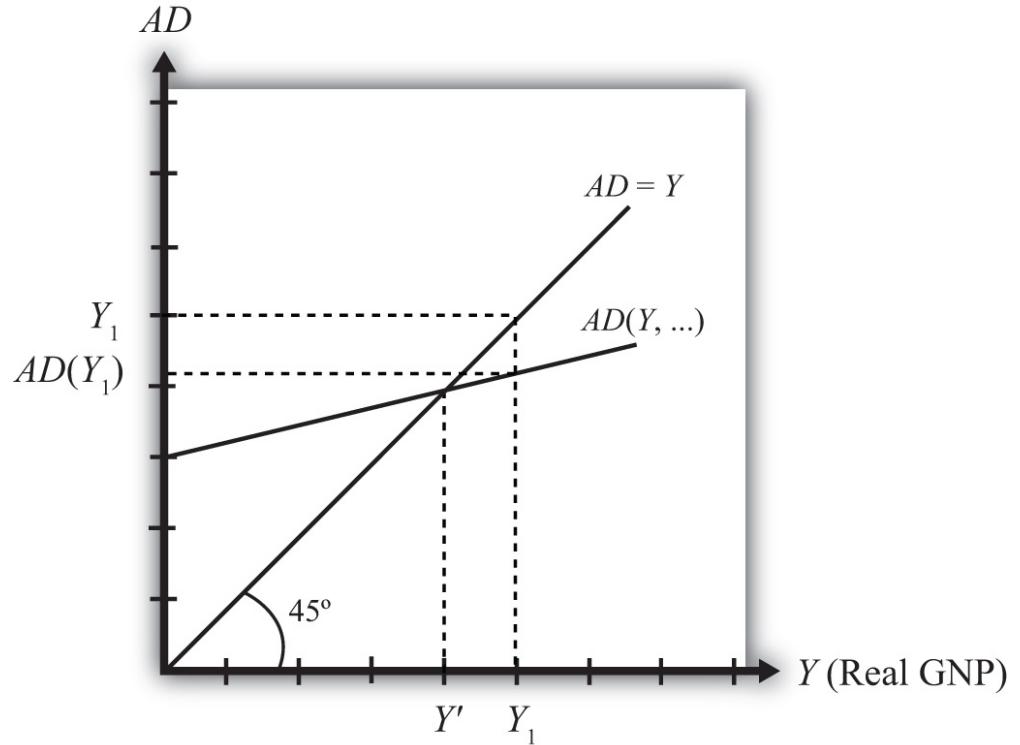
1. Learn the equilibrium stories in the G&S market that describe how GNP adjusts when it is not at its equilibrium value.

Any equilibrium in economics has an associated behavioral story to explain the forces that will move the endogenous variable to the equilibrium value. In the G&S market model, the endogenous variable is Y , real GNP. This is the variable that will change to achieve the equilibrium. Variables that do not change in the adjustment to the equilibrium are the exogenous variables. In this model, the exogenous variables are I_0 , G_0 , T , TR , $E\$/\text{£}$, $P\$_{\text{e}}$, and $P\$_{\text{f}}$. (If one uses a linear consumption demand function, the C_0 and mpc are also exogenous.) Changes in the exogenous variables are necessary to cause an adjustment to a new equilibrium. However, in telling an equilibrium story, it is typical to simply assume that the endogenous variable is not at the equilibrium (for some unstated reason) and then to explain how and why the variable will adjust to the equilibrium value.

GNP Too High

Suppose for some reason actual GNP, Y_1 , is higher than the equilibrium GNP, Y' , as shown in [Figure 19.2 "G&S Market Adjustment to Equilibrium: GNP Too High"](#). In this case, aggregate demand is read from the AD function as $AD(Y_1)$ along the vertical axis. We project aggregate supply, Y_1 , to the vertical axis using the forty-five-degree line so that we can compare supply with demand. This helps us to see that $Y_1 > AD(Y_1)$ —that is, aggregate supply is greater than aggregate demand.

Figure 19.2 G&S Market Adjustment to Equilibrium: GNP Too High



We now tell what can be called the “Inventory Story.” When total demand is less than supply, goods will begin to pile up on the shelves in stores. That’s because at current prices (and all other fixed exogenous parameters), households, businesses, and government would prefer to buy less than what is available for sale. Thus inventories begin to rise. Merchants, faced with storerooms filling up, send orders for fewer goods to producers. Producers respond to fewer orders by producing less, and thus GNP begins to fall.

As GNP falls, disposable income also falls, which causes a drop in aggregate demand as well. In the diagram, this is seen as a movement along the AD curve from Y_1 to Y' . However, GNP falls at a faster rate, along the $AD = Y$ line in the diagram. Eventually, the drop in aggregate supply catches up to the drop in demand when the equilibrium is reached at Y' . At this point, aggregate demand equals aggregate supply and there is no longer an accumulation of inventories.

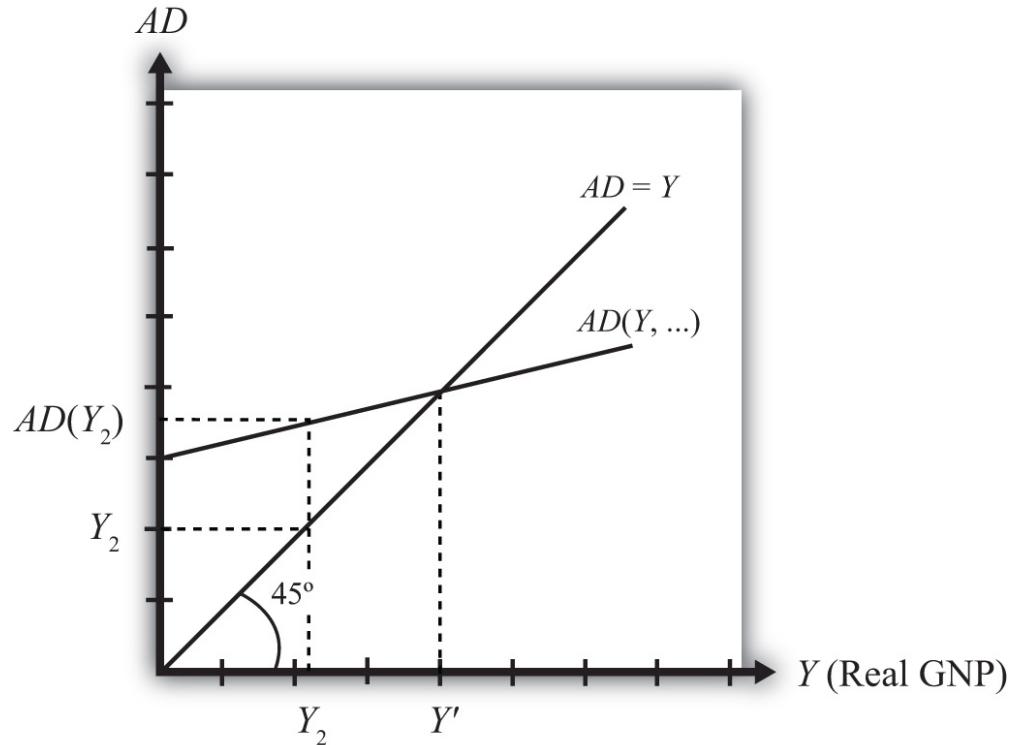
It is important to recognize a common perception or intuition that does not hold in the equilibrium adjustment process. Many students imagine a case of rising inventories and ask, “Won’t producers just lower their prices to get rid of the excess?” In real-world situations this will frequently happen; however, that response violates the *ceteris paribus* assumption of this model. We assume here

that the U.S. price level ($P_{\$}$) and consequently all prices in the economy remain fixed in the adjustment to the new equilibrium. Later, with more elaborate versions of the model, some price flexibility is considered.

GNP Too Low

Suppose for some reason, actual GNP, Y_2 , is lower than the equilibrium GNP, Y' , as shown in [Figure 19.3 "G&S Market Adjustment to Equilibrium: GNP Too Low"](#). In this case, aggregate demand is read from the AD function as $AD(Y_2)$ along the vertical axis. We project aggregate supply (Y_2) to the vertical axis using the forty-five-degree line. This shows that $AD(Y_2) > Y_2$ —that is, aggregate demand is greater than aggregate supply.

Figure 19.3 G&S Market Adjustment to Equilibrium: GNP Too Low



When total demand exceeds supply, inventories of goods that had previously been accumulated will begin to deplete in stores. That's because, at current prices (and all other fixed exogenous parameters), households, businesses, and government would prefer to buy more than is needed to keep stocks at a constant level. Merchants, faced with depleted inventories and the possibility of running out of

goods to sell, send orders to producers for greater quantities of goods. Producers respond to more orders by producing more and thus GNP begins to rise.

As GNP rises, disposable income also rises, which causes an increase in aggregate demand as well. In the diagram, this is seen as a movement along the AD curve from Y_2 to Y' . However, GNP rises at a faster rate, along the $AD = Y$ line in the diagram. Eventually, the increase in aggregate supply catches up to the increase in demand when the equilibrium is reached at Y' . At this point, aggregate demand equals aggregate supply and there is no further depletion of inventories.

KEY TAKEAWAYS

- If the actual GNP is higher than the equilibrium rate, then excess supply leads to an accumulation of inventories. Firms respond to the surplus by cutting production, causing GNP to fall until the GNP supplied is equal to aggregate demand.
- If the actual GNP is lower than the equilibrium rate, then excess demand leads to a depletion of inventories. Firms respond to the surplus by raising production, which causes GNP to rise until the GNP supplied is equal to aggregate demand.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, this will happen to store inventories when aggregate demand exceeds GNP.
 - b. Of *increase, decrease, or stay the same*, this will happen to store inventories when actual GNP is greater than equilibrium GNP.
 - c. Of *increase, decrease, or stay the same*, this is the direction of GNP change when inventories are accumulating in the Keynesian model.
 - d. Of *increase, decrease, or stay the same*, this is the direction of GNP change when inventories are depleting in the Keynesian model.
 - e. Of *faster, slower, or the same rate*, the rate of increase of aggregate demand compared to the increase in GNP as GNP rises to an equilibrium value in the Keynesian model.

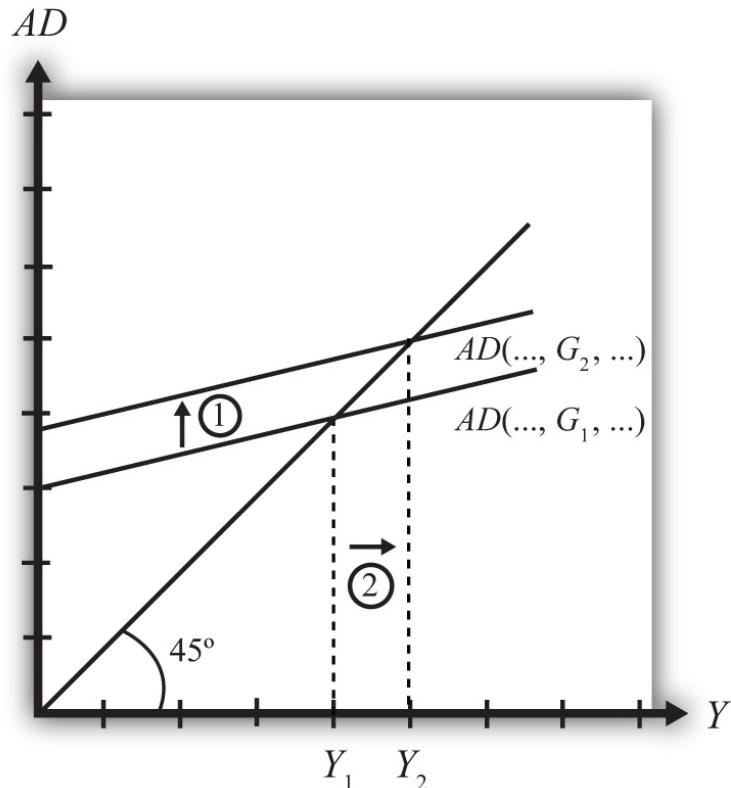
19.10 Effect of an Increase in Government Demand on Real GNP

LEARNING OBJECTIVE

1. Learn how a change in government demand affects equilibrium GNP.

Suppose the economy is initially in equilibrium in the G&S market with government demand at level G_1 and real GNP at Y_1 , shown in [Figure 19.4 "Effect of an Increase in Government Demand in the G&S Market"](#). The initial AD function is written as $AD(..., G_1, ...)$ to signify the level of government demand and to denote that other variables affect AD and are at some initial and unspecified values.

Figure 19.4 Effect of an Increase in Government Demand in the G&S Market



Next, suppose the government raises demand for G&S from G_1 to G_2 , *ceteris paribus*. The increase might arise because a new budget is passed by the legislature with new spending initiatives. The *ceteris paribus* assumption means that all other exogenous

variables are assumed to remain fixed. Most importantly in this context, this means that the increase in government demand is not paid for with increases in taxes or decreases in transfer payments.

Since higher government demand raises aggregate demand, the AD function shifts up from $AD(\dots, G_1, \dots)$ to $AD(\dots, G_2, \dots)$ (step 1). The equilibrium GNP in turn rises to Y_2 (step 2). Thus the increase in government demand causes an increase in real GNP.

The adjustment process follows the “GNP too low” story. When government demand increases, but before GNP rises to adjust, AD is greater than Y_1 . The excess demand for G&S depletes inventories, in this case for firms that supply the government, causing merchants to increase order size. This leads firms to increase output, thus raising GNP.

KEY TAKEAWAY

- In the G&S model, an increase (decrease) in government demand causes an increase (decrease) in real GNP.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the effect on equilibrium real GNP from a decrease in government demand in the G&S model.
 - b. Of *increase, decrease, or stay the same*, the effect on equilibrium real GNP caused by an increase in government demand in the G&S model.
 - c. Of *GNP too low or GNP too high*, the equilibrium story that must be told following an increase in government demand in the G&S model.
 - d. Of *GNP too low or GNP too high*, the equilibrium story that must be told following a decrease in government demand in the G&S model.
2. In the text, the effect of a change in government demand is analyzed. Use the G&S model (diagram) to individually assess the effect on equilibrium GNP caused by the following changes. Assume *ceteris paribus*.
 - a. An increase in investment demand.
 - b. An increase in transfer payments.
 - c. An increase in tax revenues.
3. Consider an economy in equilibrium in the G&S market.
 - a. Suppose investment demand decreases, *ceteris paribus*. What is the effect on equilibrium GNP?
 - b. Now suppose investment demand decreases, but *ceteris paribus* does not apply because at the same time government demand rises. What is the effect on equilibrium GNP?
 - c. In general, which of these two assumptions, *ceteris paribus* or no *ceteris paribus*, is more realistic? Explain why.
 - d. If *ceteris paribus* is less realistic, why do economic models so frequently apply the assumption?

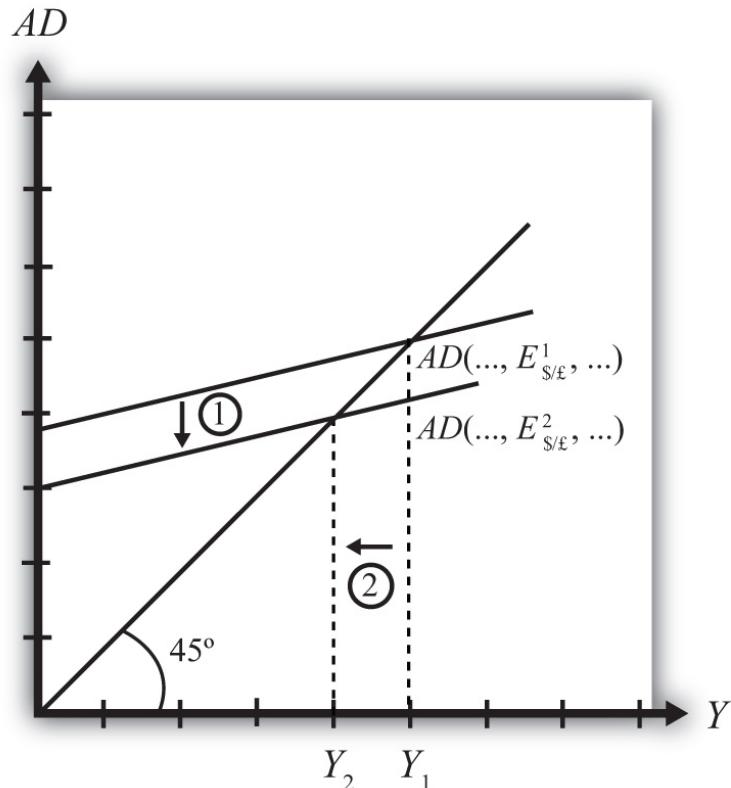
19.11 Effect of an Increase in the U.S. Dollar Value on Real GNP

LEARNING OBJECTIVE

1. Learn how a change in the U.S. dollar value affects equilibrium GNP.

Suppose the economy is initially in equilibrium in the G&S market with the exchange rate at level $E_{\$/\text{£}}^1$ and real GNP at Y_1 as shown in [Figure 19.5 "Effect of an Increase in the U.S. Dollar Value in the G&S Market"](#). The initial AD function is written as $AD(\dots, E_{\$/\text{£}}^1, \dots)$ to signify the level of the exchange rate and to denote that other variables affect AD and are at some initial and unspecified values.

Figure 19.5 Effect of an Increase in the U.S. Dollar Value in the G&S Market



Next, suppose the U.S. dollar value rises, corresponding to a decrease in the exchange rate from $E_{\$/\text{£}}^1$ to $E_{\$/\text{£}}^2$, ceteris paribus. As explained in [Chapter 19 "National Output Determination"](#), [Section 19.6 "Export and Import Demand"](#), the

increase in the spot dollar value also increases the real dollar value, causing foreign G&S to become relatively cheaper and U.S. G&S to become more expensive. This change reduces demand for U.S. exports and increases import demand, resulting in a reduction in aggregate demand. The *ceteris paribus* assumption means that all other exogenous variables are assumed to remain fixed.

Since the higher dollar value lowers aggregate demand, the AD function shifts down from $AD(\dots, E_{\$/\text{£}}^1, \dots)$ to $AD(\dots, E_{\$/\text{£}}^2, \dots)$ (step 1), and equilibrium GNP in turn falls to Y_2 (step 2). Thus the increase in the U.S. dollar value causes a decrease in real GNP.

The adjustment process follows the “GNP too high” story. When the dollar value rises but before GNP falls to adjust, $Y_1 > AD$. The excess supply of G&S raises inventories, causing merchants to decrease order size. This leads firms to decrease output, lowering GNP.

KEY TAKEAWAY

- In the G&S model, an increase (decrease) in the U.S. dollar value causes a decrease (increase) in real GNP.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the effect on equilibrium real U.S. GNP from a decrease in the value of the U.S. dollar in the G&S model.
 - b. Of *increase, decrease, or stay the same*, the effect on equilibrium real GNP caused by an increase in the value of the U.S. dollar in the G&S model.
 - c. Of *GNP too low or GNP too high*, the equilibrium story that must be told following an increase in the value of the U.S. dollar in the G&S model.
 - d. Of *GNP too low or GNP too high*, the equilibrium story that must be told following a decrease in the value of the U.S. dollar in the G&S model.
2. In the text, the effect of a change in the currency value is analyzed. Use the G&S model (diagram) to individually assess the effect on equilibrium GNP caused by the following changes. Assume *ceteris paribus*.
 - a. A decrease in the real currency value.
 - b. An increase in the domestic price level.
 - c. An increase in the foreign price level.

19.12 The J-Curve Effect

LEARNING OBJECTIVE

1. Learn about the J-curve effect that explains how current account adjustment in response to a change in the currency value will vary over time.

In the goods market model, it is assumed that the exchange rate ($E\$/\epsilon$) is directly related to current account demand in the United States. The logic of the relationship goes as follows. If the dollar depreciates, meaning $E\$/\epsilon$ rises, then foreign goods will become more expensive to U.S. residents, causing a decrease in import demand. At the same time U.S. goods will appear relatively cheaper to foreign residents, causing an increase in demand for U.S. exports. The increase in export demand and decrease in import demand both contribute to an increase in the current account demand. Since in the goods market model, any increase in demand results in an increase in supply to satisfy that demand, the dollar depreciation should also lead to an increase in the actual current account balance.

In real-world economies, however, analysis of the data suggests that in many instances a depreciating currency tends to cause, at least, a temporary increase in the deficit rather than the predicted decrease. The explanation for this temporary reversal of the cause-and-effect relationship is called the J-curve theory. In terms of future use of the AA-DD model, we will always assume the **J-curve effect**⁷ is not operating, unless otherwise specified. One should think of this effect as a possible short-term exception to the standard theory.

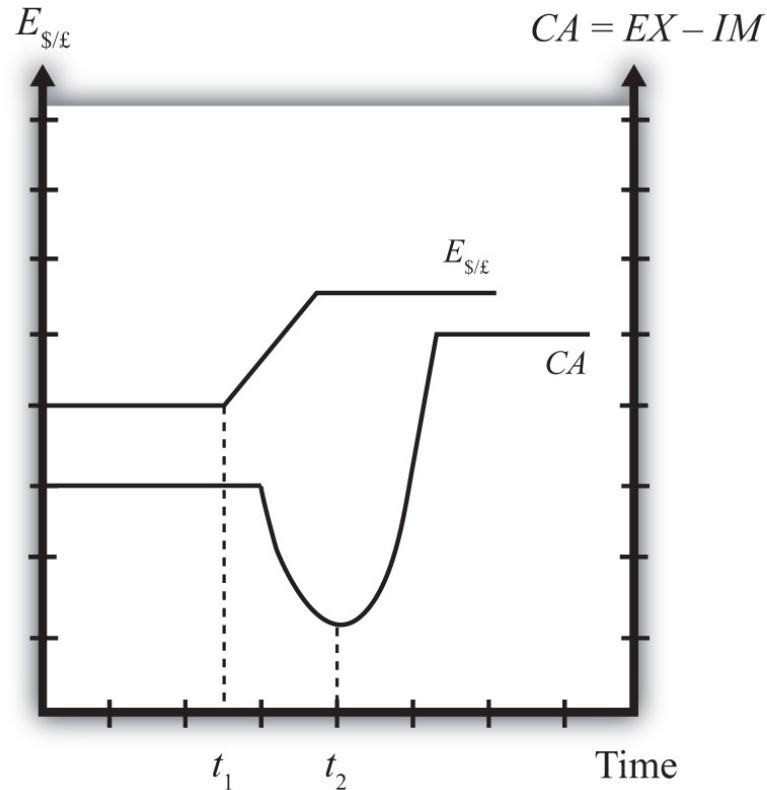
The theory of the J-curve is an explanation for the J-like temporal pattern of change in a country's trade balance in response to a sudden or substantial depreciation (or devaluation) of the currency.

Consider [Figure 19.6 "J-Curve Effect"](#), depicting two variables measured, hypothetically, over some period: the U.S. dollar / British pound ($E\$/\epsilon$) and the U.S. current account balance ($CA = EX - IM$). The exchange rate is meant to represent the average value of the dollar against all other trading country currencies and would correspond to a dollar value index that is often constructed and reported. Since the units of these two data series would be in very different scales, we imagine the exchange rate is measured along the left axis, while the CA balance is measured in

7. A theory that suggests that after a currency depreciation, the current account balance will first fall for a period of time before beginning to rise as is normally expected.

different units on the right-hand axis. With appropriately chosen scales, we can line up the two series next to each other to see whether changes in the exchange rate seem to correlate with positive or negative changes in the CA balance.

Figure 19.6 *J-Curve Effect*



As previously mentioned, the standard theory suggests a positive relationship between $E_{\$/\text{£}}$ and the U.S. current account, implying that, ceteris paribus, any dollar depreciation (an increase in $E_{\$/\text{£}}$) should cause an increase in the CA balance.

However, what sometimes happens instead, is immediately following the dollar depreciation at time t_1 , the CA balance falls for a period of time, until time t_2 is reached. In this phase, a CA deficit would become larger, not smaller.

Eventually, after period t_2 , the CA balance reverses direction and begins to increase—in other words, a trade deficit falls. The diagram demonstrates clearly how the CA balance follows the pattern of a “J” in the transition following a dollar depreciation, hence the name J-curve theory.

In the real world, the period of time thought necessary for the CA balance to traverse the J pattern is between one and two years. However, this estimate is merely a rough rule of thumb as the actual paths will be influenced by many other variable changes also occurring at the same time. Indeed, in some cases the J-curve effect may not even arise, so there is nothing automatic about it.

The reasons for the J-curve effect can be better understood by decomposing the current account balance. The basic definition of the current account is the difference between the value of exports and the value of imports. That is,

$$CA = EX - IM.$$

The current account also includes income payments and receipts and unilateral transfers, but these categories are usually small and will not play a big role in this discussion—so we'll ignore them. The main thing to take note about this definition is that the CA is measured in “value” terms, which means in terms of dollars. The way these values are determined is by multiplying the quantity of imports by the price of each imported item. We expand the CA definition by using the summation symbol and imagining summing up across all exported goods and all imported goods:

$$CA = \sum P_{EX} Q_{EX} - \sum P_{IM} Q_{IM}.$$

Here $\sum P_{EX} Q_{EX}$ represents the summation of the price times quantities of all goods exported from the country, while $\sum P_{IM} Q_{IM}$ is the summation of the price times quantities of all goods imported from the country.

However, for imported goods we could also take note that foreign products are denominated in foreign currency terms. To convert them to U.S. dollars we need to multiply by the current spot exchange rate. Thus we can expand the CA definition further by incorporating the exchange rate into the import term as follows:

$$CA = \sum P_{EX} Q_{EX} - \sum E_{\$/\text{£}} P^*_{IM} Q_{IM}.$$

Here $E_{\$/\text{£}}$ represents whatever dollar/pound rate prevailed at the time of imports, and P^*_{IM} represents the price of each imported good denominated in foreign (*) pound currency terms. Thus the value of imports is really the summation across all foreign imports of the exchange rate times the foreign price times quantity.

The J-curve theory recognizes that import and export quantities and prices are often arranged in advance and set into a contract. For example, an importer of

watches is likely to enter into a contract with the foreign watch company to import a specific quantity over some future period. The price of the watches will also be fixed by the terms of the contract. Such a contract provides assurances to the exporter that the watches he makes will be sold. It provides assurances to the importer that the price of the watches will remain fixed. Contract lengths will vary from industry to industry and firm to firm, but may extend for as long as a year or more.

The implication of contracts is that in the short run, perhaps over six to eighteen months, both the local prices and quantities of imports and exports will remain fixed for many items. However, the contracts may stagger in time—that is, they may not all be negotiated and signed at the same date in the past. This means that during any period some fraction of the contracts will expire and be renegotiated. Renegotiated contracts can adjust prices and quantities in response to changes in market conditions, such as a change in the exchange rate. Thus in the months following a dollar depreciation, contract renegotiations will gradually occur, causing eventual, but slow, changes in the prices and quantities traded.

With these ideas in mind, consider a depreciation of the dollar. In the very short run—say, during the first few weeks—most of the contract terms will remain unchanged, meaning that the prices and quantities of exports and imports will also stay fixed. The only change affecting the CA formula, then, is the increase in E/*$. Assuming all importers have not hedged their trades by entering to forward contracts, the increase in E/*$ will result in an immediate increase in the value of imports measured in dollar terms. Since the prices and quantities do not change immediately, the CA balance falls. This is what can account for the initial stage of the J-curve effect, between periods t_1 and t_2 .

As the dollar depreciation continues, and as contracts begin to be renegotiated, traders will adjust quantities demanded. Since the dollar depreciation causes imported goods to become more expensive to U.S. residents, the quantity of imported goods demanded and purchased will fall. Similarly, exported goods will appear cheaper to foreigners, and so as their contracts are renegotiated, they will begin to increase demand for U.S. exports. The changes in these quantities will both cause an increase in the current account (decrease in a trade deficit). Thus, as several months and years pass, the effects from the changes in quantities will surpass the price effect caused by the dollar depreciation and the CA balance will rise as shown in the diagram after time t_2 .

It is worth noting that the standard theory, which says that a dollar depreciation causes an increase in the current account balance, assumes that the quantity effects—that is, the effects of the depreciation on export and import demand—are

the dominant effects. The J-curve theory qualifies that effect by suggesting that although the quantity or demand effects will dominate, it may take several months or years before becoming apparent.

KEY TAKEAWAYS

- The J-curve theory represents a short-term exception to the standard assumption applied in the G&S model in which a currency depreciation causes a decrease in the trade deficit.
- The theory of the J-curve is an explanation for the J-like temporal pattern of change in a country's trade balance in response to a sudden or substantial depreciation (or devaluation) of the currency.
- The J-curve effect suggests that after a currency depreciation, the current account balance will first fall for a period of time before beginning to rise as normally expected. If a country has a trade deficit initially, the deficit will first rise and then fall in response to a currency depreciation.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *short run or long run*, the period in which the J-curve theory predicts that a country’s trade deficit will rise with a currency depreciation.
 - b. Of *short run or long run*, the period in which the J-curve theory predicts that a country’s trade deficit will fall with a currency depreciation.
 - c. Of *value of U.S. imports or quantity of U.S. imports*, this is expected to rise in the short run after a dollar depreciation according to the J-curve theory.
 - d. Of *value of Turkish imports or quantity of Turkish imports*, this is expected to fall in the long run after a Turkish lira depreciation according to the J-curve theory.
 - e. Of *increase, decrease, or stay the same*, the effect on U.S. exports in the short run due to a U.S. dollar depreciation according to the J-curve theory.
 - f. Of *increase, decrease, or stay the same*, the effect on U.S. imports in the short run due to a U.S. dollar depreciation according to the J-curve theory.

Chapter 20

The AA-DD Model

Ideally, it would be nice to develop a way to keep track of all the cause-and-effect relationships that are presumed to exist at the same time. From the previous chapters it is clear, for example, that the money supply affects the interest rates in the money market, which in turn affects the exchange rates in the foreign exchange (Forex) market, which in turn affects demand on the current account in the goods and services (G&S) market, which in turn affects the level of GNP, which in turn affects the money market, and so on. The same type of string of repercussions can be expected after many other changes that might occur. Keeping track of these effects and establishing the final equilibrium values would be a difficult task if not for a construction like the AA-DD model. This model merges the money market, the Forex market, and the G&S market into one supermodel. The construction of the AA-DD model is presented in this chapter.

20.1 Overview of the AA-DD Model

LEARNING OBJECTIVE

- Understand the basic structure and results of the AA-DD model of national output and exchange rate determination.

This chapter describes the derivation and the mechanics of the **AA-DD model**¹. The AA-DD model represents a synthesis of the three previous market models: the foreign exchange (Forex) market, the money market, and the goods and services market. In a sense, there is really very little new information presented here. Instead, the chapter provides a graphical approach to integrate the results from the three models and to show their interconnectedness. However, because so much is going on simultaneously, working with the AA-DD model can be quite challenging.

- Represents a synthesis of the foreign exchange market, the money market, and the goods and services market, showing their interconnectedness.
- The set of exchange rate and GNP combinations that maintain equilibrium in the asset markets, which is given fixed values for all other exogenous variables.
- The set of exchange rate and GNP combinations that maintain equilibrium in the goods and services market, which are given fixed values for all other exogenous variables.
- Describes the GNP level and exchange rate value at the intersection of the AA and DD curves. It represents the values that provide for equilibriums in the money market, the Forex market, and the G&S market simultaneously.

The AA-DD model is described with a diagram consisting of two curves (or lines): an **AA curve**² representing asset market equilibriums derived from the money market and foreign exchange markets and a **DD curve**³ representing goods market (or demand) equilibriums. The intersection of the two curves identifies a market equilibrium in which each of the three markets is simultaneously in equilibrium. Thus we refer to this equilibrium as a **superequilibrium**⁴.

Results

The main results of this section are descriptive and purely mechanical. The chapter describes the derivation of the AA and DD curves, explains how changes in exogenous variables will cause shifts in the curves, and explains adjustment from one equilibrium to another.

- The DD curve is the set of exchange rate and GNP combinations that maintain equilibrium in the goods and services market, given fixed values for all other exogenous variables.
- The DD curve shifts rightward whenever government demand (G), investment demand (I), transfer payments (TR), or foreign prices (P_E) increase or when taxes (T) or domestic prices (P_S) decrease. Changes in the opposite direction cause a leftward shift.

- c. The AA curve is the set of exchange rate and GNP combinations that maintain equilibrium in the asset markets, given fixed values for all other exogenous variables.
- d. The AA curve shifts upward whenever money supply (M^S), foreign interest rates (i_E), or the expected exchange rate ($E_{\$/E}^e$) increase or when domestic prices ($P_{\$}$) decrease. Changes in the opposite direction cause a downward shift.
- e. The intersection of the AA and DD curves depicts a superequilibrium in an economy since at that point the goods and services market, the domestic money market, and the foreign exchange market are all in equilibrium simultaneously.
- f. Changes in any exogenous variable that is not plotted on the axes (anything but Y and $E_{\$/E}$) will cause a shift of the AA or DD curves and move the economy out of equilibrium, temporarily. Adjustment to a new equilibrium follows the principle that adjustment in the asset markets occurs much more rapidly than adjustment in the goods and services market. Thus adjustment to the AA curve will always occur before adjustment to the DD curve.

Connections

The AA-DD model will allow us to understand how changes in macroeconomic policy—both monetary and fiscal—can affect key aggregate economic variables when a country is open to international trade and financial flows while accounting for the interaction of the variables among themselves. Specifically, the model is used to identify potential effects of fiscal and monetary policy on exchange rates, trade balances, GDP levels, interest rates, and price levels both domestically and abroad. In subsequent chapters, analyses will be done under both floating and fixed exchange rate regimes.

KEY TAKEAWAYS

- The AA-DD model integrates the workings of the money-Forex market and the G&S model into one supermodel.
- The AA curve is derived from the money-Forex model. The DD curve is derived from the G&S model.
- The intersection of the AA and DD curves determines the equilibrium values for real GNP and the exchange rate.
- Comparative statics exercises using the AA-DD model allow one to identify the effects of changes in exogenous variables on the level of GDP and the exchange rate, while assuring that the Forex, the money market, and the G&S market all achieve simultaneous equilibrium.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. At the intersection of the AA and DD curves, the goods and services market, the money market, and this market are simultaneously in equilibrium.
 - b. The term used to describe the type of equilibrium at the intersection of the AA curve and the DD curve.

20.2 Derivation of the DD Curve

LEARNING OBJECTIVE

1. Learn how to derive the DD curve from the G&S model.

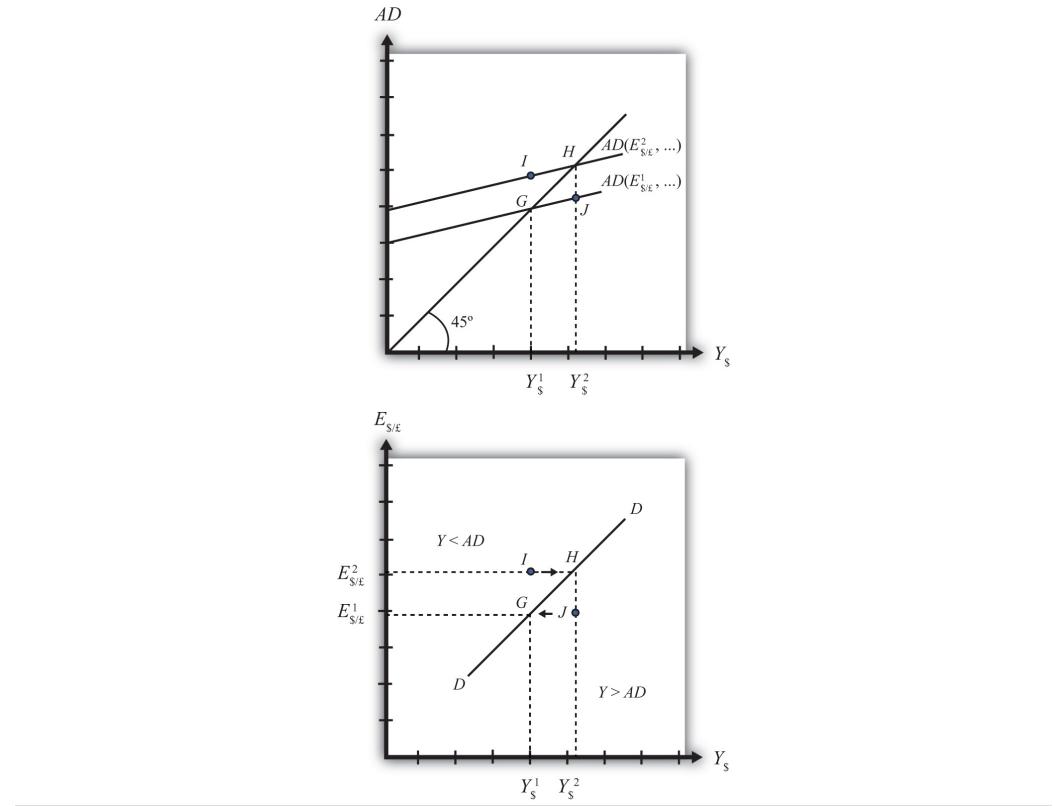
The DD curve is derived by transferring information described in the goods and services (G&S) market model onto a new diagram to show the relationship between the exchange rate and equilibrium gross national product (GNP). The original G&S market, depicted in the top part of [Figure 20.1 "Derivation of the DD Curve"](#), plots the aggregate demand (AD) function with respect to changes in U.S. GNP ($Y_{\$}$).

Aggregate demand is measured along the vertical axis and aggregate supply (or the GNP) is measured on the horizontal axis. As discussed in [Chapter 19 "National Output Determination"](#), [Section 19.7 "The Aggregate Demand Function"](#), the AD function is dependent upon several different exogenous variables, most notably the exchange rate between domestic and foreign currency ($E_{\$/\text{£}}$). However, AD is also affected by investment demand (I), government demand (G), government tax revenues (T), government transfer payments (TR), and the price level in the domestic ($P_{\$}$) and foreign ($P_{\text{£}}$) countries. The endogenous variable in the model is U.S. GNP ($Y_{\$}$). (See [Table 20.1 "G&S Market"](#) for a quick reference.) In this exercise, since our focus is on the exchange rate, we label the AD function in [Figure 20.1 "Derivation of the DD Curve"](#) as $AD(E_{\$/\text{£}}, \dots)$, where the ellipsis (...) is meant to indicate there are other unspecified variables that also influence AD.

Table 20.1 G&S Market

Exogenous Variables	$E_{\$/\text{£}}, I, G, T, TR, P_{\$}, P_{\text{£}}$
Endogenous Variable	$Y_{\$}$

Figure 20.1 Derivation of the DD Curve



Initially, let's assume the exchange rate is at a value in the market given by $E_{\$/\epsilon}^1$ ¹. We need to remember that all the other variables that affect AD are also at some initial level. Written explicitly, we could write AD as $AD(E_{\$/\epsilon}^1, l^1, G^1, T^1, TR^1, P_{\$}^1, P_{\epsilon}^1)$. The AD function with exchange rate $E_{\$/\epsilon}^1$ intersects the forty-five-degree line at point G which determines the equilibrium level of GNP given by $Y_{\1 . These two values are transferred to the lower diagram at point G determining one point on the DD curve ($Y_{\$}^1, E_{\$/\epsilon}^1$).

Next, suppose $E_{\$/\epsilon}$ rises from $E_{\$/\epsilon}^1$ to $E_{\$/\epsilon}^2$, ceteris paribus. This corresponds to a depreciation of the U.S. dollar with respect to the British pound. The ceteris paribus assumption means that investment, government, taxes, and so on stay fixed at levels I^1, G^1, T^1 , and so on. Since a dollar depreciation makes foreign G&S relatively more expensive and domestic goods relatively cheaper, AD shifts up to $AD(E_{\$/\epsilon}^2, \dots)$. The equilibrium shifts to point H at a GNP level $Y_{\2 . These two values are transferred to the lower diagram at point H, determining a second point on the DD curve ($Y_{\$}^2, E_{\$/\epsilon}^2$).

The line drawn through points *G* and *H* on the lower diagram is called the DD curve. The DD curve plots an equilibrium GNP level for every possible exchange rate that may prevail, ceteris paribus. Stated differently, the DD curve is the combination of exchange rates and GNP levels that maintain equilibrium in the G&S market, ceteris paribus. We can think of it as the set of aggregate demand equilibriums.

A Note about Equilibriums

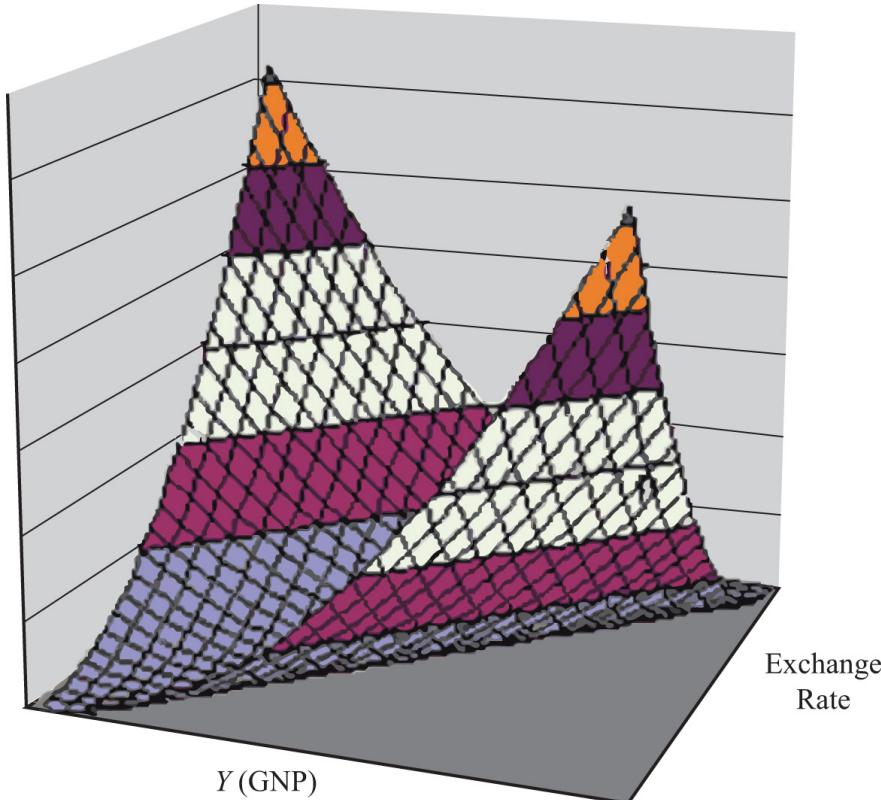
An equilibrium in an economic model typically corresponds to a point toward which the endogenous variable values will converge based on some behavioral assumption about the participants in the model. In this case, equilibrium is not represented by a single point. Instead every point along the DD curve is an equilibrium value.

If the economy were at a point above the DD curve, say, at *I* in the lower diagram, the exchange rate would be $E_{\$/\text{£}}^2$ and the GNP level at $Y_{\1 . This corresponds to point *I* in the upper diagram where $AD > Y$, read off the vertical axis. In the G&S model, whenever aggregate demand exceeds aggregate supply, producers respond by increasing supply, causing GNP to rise. This continues until $AD = Y$ at point *H*. For all points to the left of the DD curve, $AD > Y$, therefore the behavior of producers would cause a shift to the right from any point like *I* to a point like *H* on the DD curve.

Similarly, at a point such as *J*, to the right of the DD curve, the exchange rate is $E_{\$/\text{£}}^1$ and the GNP level is at $Y_{\2 . This corresponds to point *J* in the upper diagram above where aggregate demand is less than supply ($AD < Y$). In the G&S model, whenever supply exceeds demand, producers respond by reducing supply, thus GNP falls. This continues until $AD = Y$ at point *G*. For all points to the right of the DD curve, $AD < Y$, therefore the behavior of producers would cause a shift to the left from any point like *J* to a point like *G* on the DD curve.

A useful analogy is to think of the DD curve as a river flowing through a valley. (See the 3-D diagram in [Figure 20.2 "A 3-D DD Curve"](#).) The hills rise up to the right and left along the upward-sloping DD curve. Just as gravity will move a drop of water downhill onto the river valley, firm behavior will move GNP much in the same way: right or left to the lowest point along the DD curve.

Figure 20.2 A 3-D DD Curve



KEY TAKEAWAYS

- The DD curve plots an equilibrium GNP level for every possible exchange rate that may prevail, *ceteris paribus*.
- Every point on a DD curve represents an equilibrium value in the G&S market.
- The DD curve is positively sloped because an increase in the exchange rate (meaning a decrease in the U.S. dollar value) raises equilibrium GNP in the G&S model.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This is what has happened to its currency value if an economy’s exchange rate and GNP combination moves upward along an upward-sloping DD curve.
 - b. Of *greater than, less than, or equal to*, this is how aggregate demand compares to GNP when the economy has an exchange rate and GNP combination that places it to the left of the DD curve.
 - c. Of *greater than, less than, or equal to*, this is how aggregate demand compares to GNP when the economy has an exchange rate and GNP combination that places it on the DD curve.
 - d. The equilibriums along a DD curve satisfy this condition.

20.3 Shifting the DD Curve

LEARNING OBJECTIVE

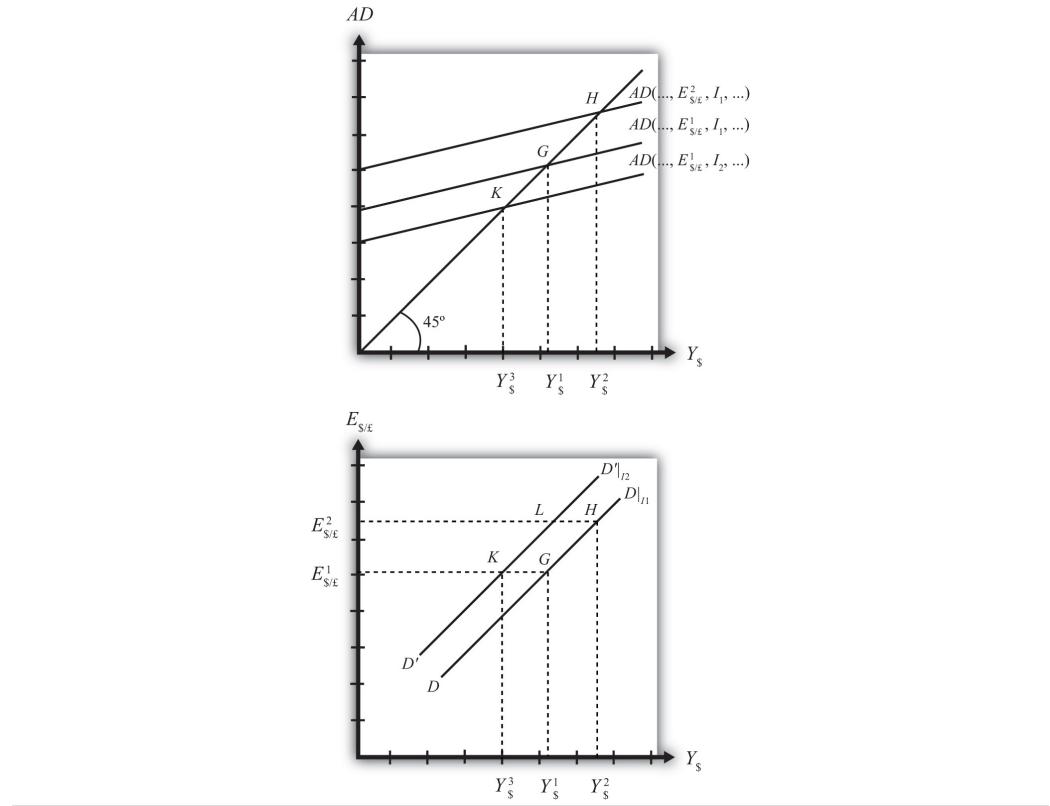
1. Learn which exogenous variables will shift the DD curve and in which direction.

The DD curve depicts the relationship between changes in one exogenous variable and one endogenous variable within the goods and services (G&S) market model. The exogenous variable assumed to change is the exchange rate. The endogenous variable affected is the gross national product (GNP). At all points along the DD curve, it is assumed that all other exogenous variables remain fixed at their original values.

The DD curve will shift, however, if there is a change in any of the other exogenous variables. We illustrate how this works in [Figure 20.3 "DD Curve Effects from a Decrease in Investment Demand"](#). Here, we assume that the level of investment demand in the economy falls from its initial level I_1 to a lower level I_2 .

At the initial investment level (I_1) and initial exchange rate ($E_{\$/\text{£}}^1$) the AD curve is given by $AD(\dots, E_{\$/\text{£}}^1, I_1, \dots)$. The AD curve intersects the forty-five-degree line at point G, which is transferred to point G on the DD curve below. If the investment level and all other exogenous variables remain fixed while the exchange rate increases to $E_{\$/\text{£}}^2$, then the AD curve shifts up to $AD(\dots, E_{\$/\text{£}}^2, I_1, \dots)$, generating the equilibrium points H in both diagrams. This exercise plots out the initial DD curve labeled $DD|_{I_1}$ in the lower diagram connecting points G and H. $DD|_{I_1}$ is read as “the DD curve given that $I = I_1$.”

Figure 20.3 DD Curve Effects from a Decrease in Investment Demand



Now, suppose I falls to I_2 . The reduction in I leads to a reduction in AD , *ceteris paribus*. At the exchange rate $E_{\$/\epsilon}^1$, the AD curve will shift down to $AD(\dots, E_{\$/\epsilon}^1, I_2, \dots)$, intersecting the forty-five-degree line at point K . Point K above, which corresponds to the combination $(E_{\$/\epsilon}^1, I_2)$, is transferred to point K on the lower diagram. This point lies on a new DD curve because a second exogenous variable, namely I , has changed. If we maintain the investment level at I_2 and change the exchange rate up to $E_{\$/\epsilon}^2$, the equilibrium will shift to point L (shown only on the lower diagram), plotting out a whole new DD curve. This DD curve is labeled $D'D'|_{I_2}$, which means “the DD curve given is $I = I_2$.”

The effect of a decrease in investment demand is to lower aggregate demand and shift the DD curve to the left. Indeed, a change in any exogenous variable that reduces aggregate demand, except the exchange rate, will cause the DD curve to shift to the left. Likewise, any change in an exogenous variable that causes an increase in aggregate demand will cause the DD curve to shift right. An exchange rate change will not shift DD because its effect is accounted for by the DD curve itself. Note that curves or lines can shift only when a variable that is *not* plotted on the axis changes.

The following table presents a list of all variables that can shift the DD curve right and left. The up arrow indicates an increase in the variable, and the down arrow indicates a decrease.

DD right-shifters	$\uparrow G \uparrow I \downarrow T \uparrow TR \downarrow P_{\$} \uparrow P_{\text{£}}$
DD left-shifters	$\downarrow G \downarrow I \uparrow T \downarrow TR \uparrow P_{\$} \downarrow P_{\text{£}}$

Refer to [Chapter 19 "National Output Determination"](#) for a complete description of how and why each variable affects aggregate demand. For easy reference, recall that G is government demand, I is investment demand, T refers to tax revenues, TR is government transfer payments, $P_{\$}$ is the U.S. price level, and $P_{\text{£}}$ is the foreign British price level.

KEY TAKEAWAYS

- The effect of an increase in investment demand (an increase in government demand, a decrease in taxes, an increase in transfer payments, a decrease in U.S. prices, or an increase in foreign prices) is to raise aggregate demand and shift the DD curve to the right.
- The effect of a decrease in investment demand (a decrease in government demand, an increase in taxes, a decrease in transfer payments, an increase in U.S. prices, or a decrease in foreign prices) is to lower aggregate demand and shift the DD curve to the left.

EXERCISE

1. Identify whether the DD curve shifts in response to each of the following changes. Indicate whether the curve shifts up, down, left, or right. Possible answers are DD right, DD left, or neither.
 - a. Decrease in government transfer payments.
 - b. Decrease in the foreign price level.
 - c. Increase in foreign interest rates.
 - d. Decrease in the expected exchange rate $E_{\$/\text{£}}^e$.
 - e. Decrease in U.S. GNP.
 - f. Decrease in the U.S. money supply.

20.4 Derivation of the AA Curve

LEARNING OBJECTIVE

1. Learn how to derive the AA curve from the money-Forex model.

The AA curve is derived by transferring information described in the money market and foreign exchange market models onto a new diagram to show the relationship between the exchange rate and equilibrium GNP. (At this point we will substitute GNP for its virtually equivalent measure, GDP, as a determinant of real money demand.) Since both models describe supply and demand for money, which is an asset, I'll refer to the two markets together as the asset market. The foreign exchange market, depicted in the top part of [Figure 20.4 "Derivation of the AA Curve"](#), plots the rates of return on domestic U.S. assets ($RoR_{\$}$) and foreign British assets (RoR_E). (See [Chapter 16 "Interest Rate Parity"](#), [Section 16.3 "Forex Equilibrium with the Rate of Return Diagram"](#) for a complete description.) The domestic U.S. money market, in the lower quadrant, plots the real U.S. money supply ($M_{\$}^S/P_{\$}$) and real money demand ($L(i_{\$}, Y_{\$})$). The asset market equilibriums have several exogenous variables that determine the positions of the curves and the outcome of the model. These exogenous variables are the foreign British interest rate (i_E) and the expected future exchange rate ($E_{\$/\e), which influence the foreign British rate of return (RoR_E); the U.S. money supply ($M_{\S) and domestic U.S. price level ($P_{\$}$), which influence real money supply; and U.S. GNP ($Y_{\$}$), which influences real money demand. The endogenous variables in the asset model are the domestic interest rates ($i_{\$}$) and the exchange rate ($E_{\$/\$}$). See [Table 20.2 "Asset Market \(Money + Forex\)"](#) for easy reference.

Figure 20.4 Derivation of the AA Curve

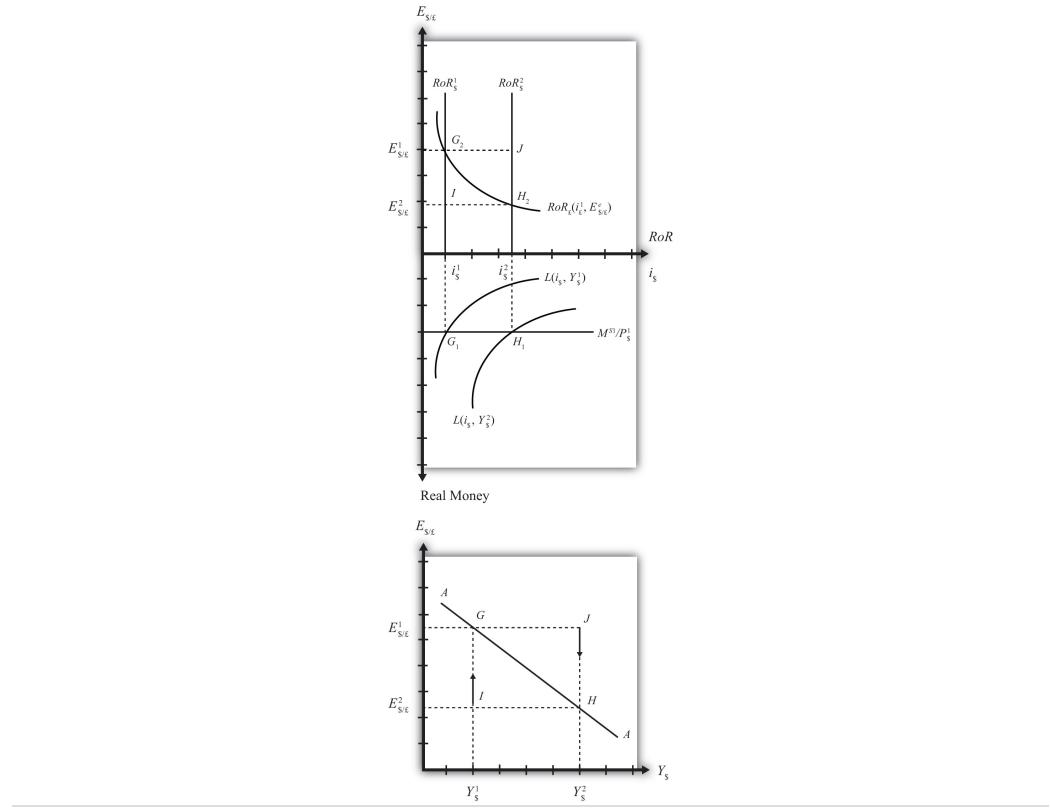


Table 20.2 Asset Market (Money + Forex)

Exogenous Variables	$i_E, E_{\$/\text{£}}^e, M_{\$}^S, P_{\$}, Y_{\$}$
Endogenous Variables	$i_{\$}, E_{\$/\text{£}}$

Initially, let's assume GNP is at a value in the market given by $Y_{\1 . We need to remember that all the other exogenous variables that affect the asset market are also at some initial level such as $i_E^1, E_{\$/\text{£}}^{e1}, M_{\S1 , and $P_{\1 . The real money demand function with GNP level $Y_{\1 intersects with real money supply at point G_1 in the money market diagram determining the interest rate $i_{\1 . The interest rate in turn determines $RoR_{\1 , which intersects with RoR_E at point G_2 , determining the equilibrium exchange rate $E_{\$/\text{£}}^{e1}$. These two values are transferred to the lowest diagram at point G , establishing one point on the AA curve ($Y_{\$}^1, E_{\$/\text{£}}^{e1}$).

Next, suppose GNP rises, for some unstated reason, from $Y_{\1 to $Y_{\2 , ceteris paribus. The ceteris paribus assumption means that all exogenous variables in the model

remain fixed. Since the increase in GNP raises real money demand, $L(i\$, Y\$)$, it shifts out to $L(i\$, Y\$^2)$. The equilibrium shifts to point H_1 , raising the equilibrium interest rate to $i\2 . The $RoR\$\text{}$ line shifts right with the interest rate, determining a new equilibrium in the Forex at point H_2 with equilibrium exchange rate $E\$/\epsilon^2$. These two values are then transferred to the diagram below at point H , establishing a second point on the AA curve $(Y\$^2, E\$/\epsilon^2)$.

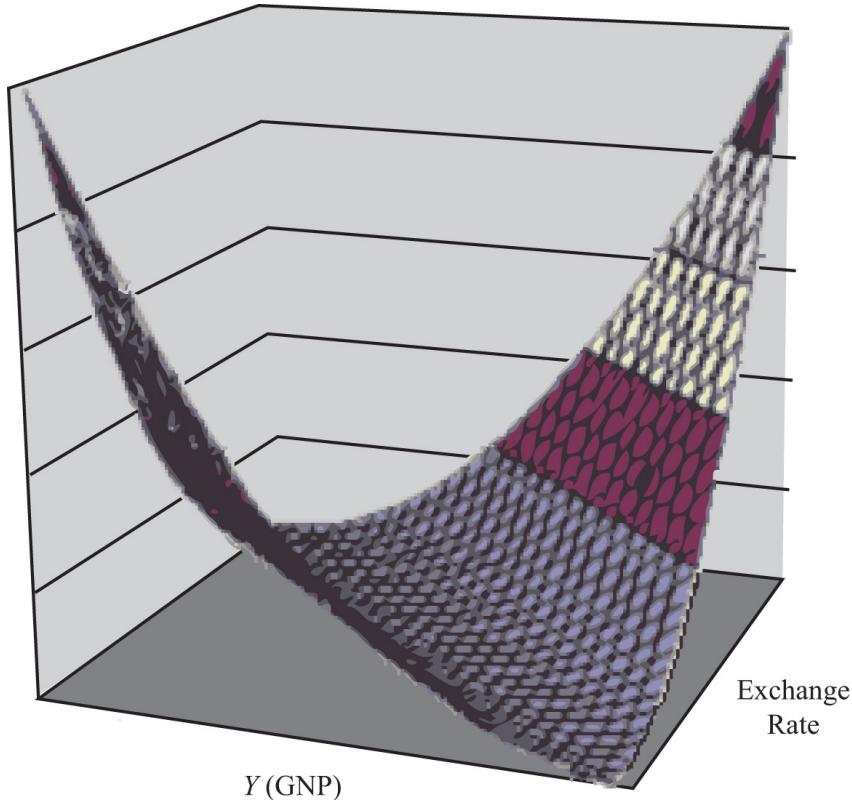
The line drawn through points G and H on the lower diagram in [Figure 20.4 "Derivation of the AA Curve"](#) is called the AA curve. The AA curve plots an equilibrium exchange rate for every possible GNP level that may prevail, ceteris paribus. Stated differently, the AA curve is the combination of exchange rates and GNP levels that maintain equilibrium in the asset market, ceteris paribus. We can think of it as the set of aggregate asset equilibria.

A Note about Equilibria

If the economy were at a point off the AA curve, like at I in the lower diagram, the GNP level is at $Y\1 and the exchange rate is $E\$/\epsilon^1$. This corresponds to point I in the upper diagram where $RoR_E > RoR\$\text{}$. In the Forex model, when foreign assets have a higher rate of return than domestic assets, investors respond by buying pounds in exchange for dollars in the foreign exchange market. This leads to a depreciation of the dollar and an increase in $E\$/\epsilon$. This continues until $RoR_E = RoR\$\text{}$ at point G . For all points below the AA curve, $RoR_E > RoR\$\text{}$; therefore, the behavior of investors would cause an upward adjustment toward the AA curve from any point like I to a point like G .

Similarly, at a point such as J , above the DD curve, the GNP level is at Y_2 and the exchange rate is $E\$/\epsilon^1$. This corresponds to point J in the upper diagram where $RoR\$_{} > RoR_E$ and the rate of return on dollar assets is greater than the rate of return abroad. In the Forex model, when U.S. assets have a higher rate of return than foreign assets, investors respond by buying dollars in exchange for pounds in the foreign exchange market. This leads to an appreciation of the dollar and a decrease in $E\$/\epsilon$. This continues until $RoR_E = RoR\$\text{}$ at point H . For all points above the AA curve, $RoR\$_{} > RoR_E$; therefore, the behavior of investors would cause a downward adjustment to the AA curve from a point like J to a point like H .

Figure 20.5 A 3-D AA Curve



As with the DD curve, it is useful to think of the AA curve as a river flowing through a valley. (See the 3-D diagram in [Figure 20.5 "A 3-D AA Curve"](#).) The hills rise up both above and below. Just as gravity will move a drop of water down the hill to the river valley, in much the same way, investor behavior will move the exchange rate up or down to the lowest point lying on the AA curve.

KEY TAKEAWAYS

- The AA curve plots an equilibrium exchange rate level for every possible GNP value that may prevail, *ceteris paribus*.
- Every point on an AA curve represents an equilibrium value in the money-Forex market.
- The AA curve is negatively sloped because an increase in the real GNP lowers the equilibrium exchange rate in the money-Forex model.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This is what has happened to its GNP if an economy’s exchange rate and GNP combination move downward along a downward-sloping AA curve.
 - b. Of *greater than, less than, or equal to*, this is how the rate of return on domestic assets compares to the rate of return on foreign assets when the economy has an exchange rate and GNP combination that places it above the AA curve.
 - c. Of *greater than, less than, or equal to*, this is how the rate of return on domestic assets compares to the rate of return on foreign assets when the economy has an exchange rate and GNP combination that places it on the AA curve.
 - d. The equilibriums along an AA curve satisfy this condition.

20.5 Shifting the AA Curve

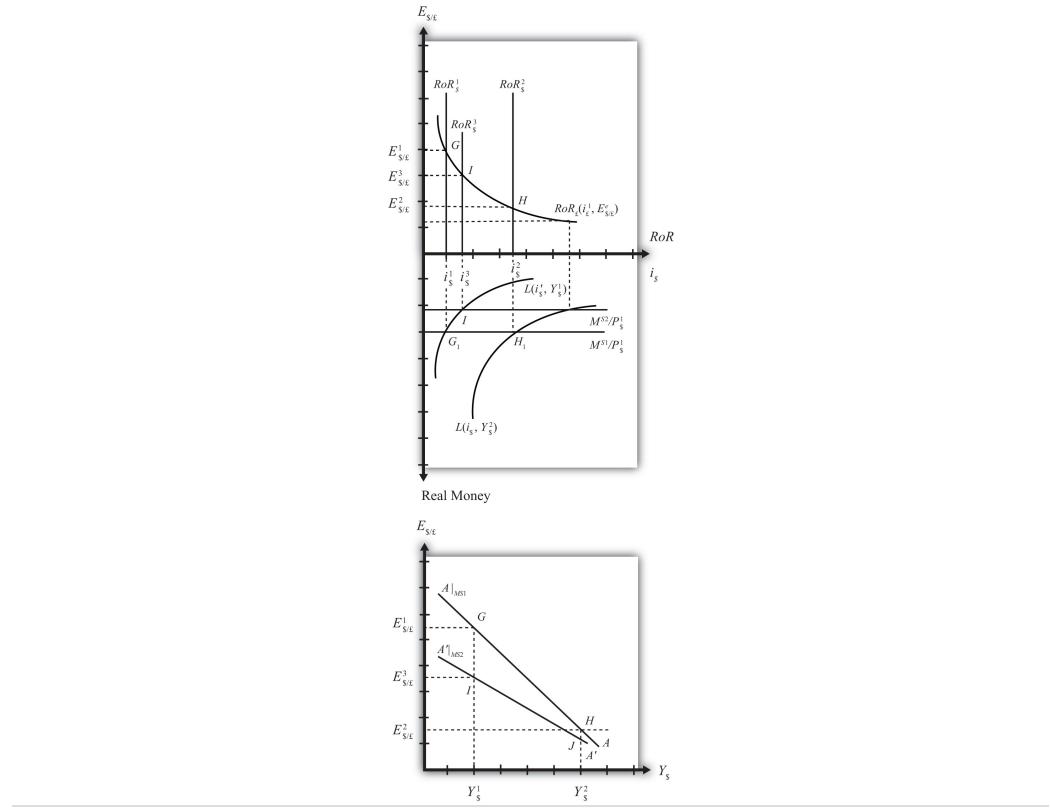
LEARNING OBJECTIVE

1. Learn which exogenous variables will shift the AA curve and in which direction.

The AA curve depicts the relationship between changes in one exogenous variable and one endogenous variable within the asset market model. The exogenous variable changed is gross national product (GNP). The endogenous variable affected is the exchange rate. At all points along the AA curve, it is assumed that all other exogenous variables remain fixed at their original values.

The AA curve will shift if there is a change in any of the other exogenous variables. We illustrate how this works in [Figure 20.6 "AA Curve Effects from a Decrease in the Money Supply"](#), where we assume that the money supply in the economy falls from its initial level M^{S1} to a lower level M^{S2} .

Figure 20.6 AA Curve Effects from a Decrease in the Money Supply



At the initial money supply (M^S1) and initial GNP level $Y_\1 , real money demand intersects real money supply at point G , determining the interest rate $i_\1 . This in turn determines the rate of return on U.S. assets, $RoR_\1 , which intersects the foreign British RoR_F at G in the upper diagram, determining the equilibrium exchange rate $E_\$/\text{\textsterling}^1$. If the money supply and all other exogenous variables remain fixed, while GNP increases to $Y_\2 , the equilibriums shift to points H in the lower and upper diagrams, determining exchange rate $E_\$/\text{\textsterling}^2$. This exercise plots out the initial AA curve labeled $AA|_{MS1}$ in the lower diagram connecting points G and H . Note, $AA|_{MS1}$ is read as “the AA curve given that $M^S = M^S1$.”

Now, suppose the money supply M^S falls to M^{S2} . The reduction in M^S leads to a reduction in the real money supply, which, at GNP level $Y_\1 , shifts the money market equilibrium to point I , determining a new interest rate, $i_\3 . In the Forex market, the rate of return rises to $RoR_\3 , which determines the equilibrium exchange rate $E_\$/\text{\textsterling}^3$. The equilibriums at points I corresponding to the combination $(Y_\$^1, E_\$/\text{\textsterling}^3)$ are transferred to point I on the lower diagram. This point lies on a new AA curve because a second exogenous variable, namely, M^S , has changed. If we

maintain the money supply at M^S and change the GNP up to $Y_{\2 , the equilibrium will shift to point J (shown only on the lower diagram), plotting out a whole new AA curve. This AA curve is labeled $A'A'|_{MS2}$, which means “the AA curve given that $M^S = M^{S2}$.”

The effect of a decrease in the money supply is to shift the AA curve downward. Indeed, a change in any exogenous variable in the asset markets that reduces the equilibrium exchange rate, except a change in GNP, will cause the AA curve to shift down. Likewise, any change in an exogenous variable that causes an increase in the exchange rate will cause the AA curve to shift up. A change in GNP will *not* shift AA because its effect is accounted for by the AA curve itself. Note that curves or lines can shift only when a variable *not* plotted on the axis changes.

The following table presents a list of all variables that can shift the AA curve up and down. The up arrow indicates an increase in the variable, and a down arrow indicates a decrease.

AA up-shifters	$\uparrow M^S \downarrow P_{\$} \uparrow i_E \uparrow E_{\$/E}^e$
AA down-shifters	$\downarrow M^S \uparrow P_{\$} \downarrow i_E \downarrow E_{\$/E}^e$

Refer to [Chapter 16 "Interest Rate Parity"](#) and [Chapter 18 "Interest Rate Determination"](#) for a complete description of how and why each variable affects the exchange rate. For easy reference though, recall that M^S is the U.S. money supply, $P_{\$}$ is the U.S. price level, i_E is the foreign British interest rate, and $E_{\$/E}^e$ is the expected future exchange rate.

KEY TAKEAWAYS

- The effect of an increase in the money supply (or a decrease in the price level, an increase in foreign interest rates, or an increase in the expected exchange rate [as defined]) is to shift the AA curve upward.
- The effect of a decrease in the money supply (or an increase in the price level, a decrease in foreign interest rates, or a decrease in the expected exchange rate [as defined]) is to shift the AA curve downward.

EXERCISE

1. Identify whether the AA curve shifts in response to each of the following changes. Indicate whether the curve shifts up, down, left, or right. Possible answers are AA right, AA left, or neither.
 - a. Decrease in government transfer payments.
 - b. Decrease in the foreign price level.
 - c. Increase in foreign interest rates.
 - d. Decrease in the expected exchange rate $E_{\$/\text{£}}^e$.
 - e. Decrease in U.S. GNP.
 - f. Decrease in the U.S. money supply.

20.6 Superequilibrium: Combining DD and AA

LEARNING OBJECTIVE

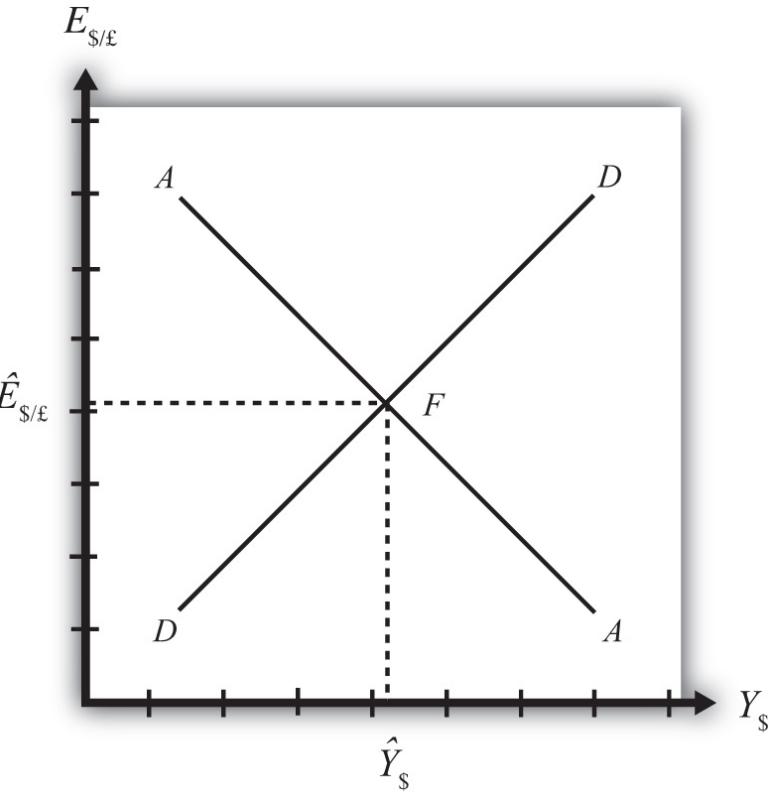
1. Apply the AA curve and the DD curve to define a superequilibrium.

The DD curve represents the set of equilibriums in the goods and services (G&S) market. It describes an equilibrium gross national product (GNP) level for each and every exchange rate that may prevail. Due to the assumption that firms respond to excess demand by increasing supply (and to excess supply by decreasing supply), GNP rises or falls until the economy is in equilibrium on the DD curve.

The AA curve represents the set of equilibriums in the asset market. It indicates an equilibrium exchange rate for each and every GNP level that might prevail. Due to the assumption that investors will demand foreign currency when the foreign rate of return exceeds the domestic return and that they will supply foreign currency when the domestic rate of return exceeds the foreign return, the exchange rate will rise or fall until the economy is in equilibrium on the AA curve.

Since both the G&S market and the asset markets are operating concurrently, equilibriums in both markets can only occur where the DD curve intersects the AA curve. This is shown in [Figure 20.7 "AA-DD Superequilibrium"](#) at point *F*, with equilibrium GNP ($\underline{\$}$) and exchange rate ($\hat{E}_{\$/£}$). It is worth emphasizing that at point *F*, the three markets—that is, the G&S market, the money market, and the foreign exchange market—are in equilibrium simultaneously. For this reason, point *F* is more than a plain old equilibrium; instead it is a *superequilibrium*.

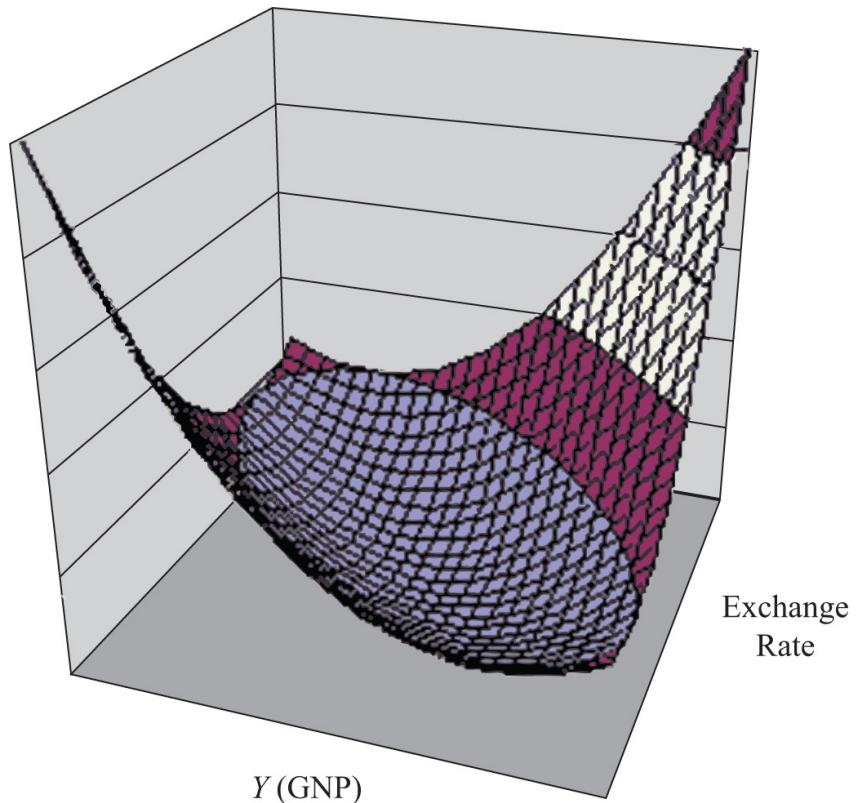
Figure 20.7 AA-DD Superequilibrium



The superequilibrium point is where we would expect behavioral responses by firms, households, and investors to move the exchange rate and GNP level, assuming the exogenous variables remain fixed at their original levels and assuming sufficient time is allowed for adjustment to the equilibrium to take place.

The equilibrium at F is like the lowest point of two intersecting valleys that reach their combined lowest point at a pool where the two valleys meet. A 3-D rendition of this is shown in [Figure 20.8 "A 3-D AA-DD Depiction"](#). The steepness of the valleys is meant to represent the speed of adjustment. Thus the AA valley is drawn much steeper than the DD valley to reflect the much more rapid adjustment in the asset markets in comparison to goods market adjustment. Anytime the economy is away from the equilibrium, forces will act to move it to the pool in the center. However, as will be shown later, adjustment to the AA curve will occur much faster than adjustment to the DD curve.

Figure 20.8 A 3-D AA-DD Depiction



KEY TAKEAWAY

- A superequilibrium describes the GNP level and exchange rate value at the intersection of the AA and DD curves. It represents the values that provide for equilibriums in the money market, the Forex market, and the G&S market simultaneously.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This market is in equilibrium when an economy is on an AA curve.
 - b. This market is in equilibrium when an economy is on a DD curve.
 - c. The term used to describe the equilibrium at the intersection of a DD curve and an AA curve.

20.7 Adjustment to the Superequilibrium

LEARNING OBJECTIVE

1. Learn how to describe the complete adjustment to equilibrium in the AA-DD model.

In order to discuss adjustment to the superequilibrium, we must first talk about how an economy can end up out of equilibrium. This will occur anytime there is a change in one or more of the exogenous variables that cause the AA or DD curves to shift. In a real economy, we should expect these variables to be changing frequently. Variables such as interest rates will certainly change every day. A variable such as the average expected future exchange rate held by investors probably changes every minute. Each time an exogenous variable changes, the superequilibrium point will shift, setting off behavioral responses by households, businesses, and investors that will affect the exchange rate and gross national product (GNP) in the direction of the new superequilibrium. However, as we will indicate below, the adjustment process will take time, perhaps several months or more, depending on the size of the change. Since we should expect that as adjustment to one variable change is in process, other exogenous variables will also change, we must recognize that the superequilibrium is really like a moving target. Each day, maybe each hour, the target moves, resulting in a continual adjustment process.

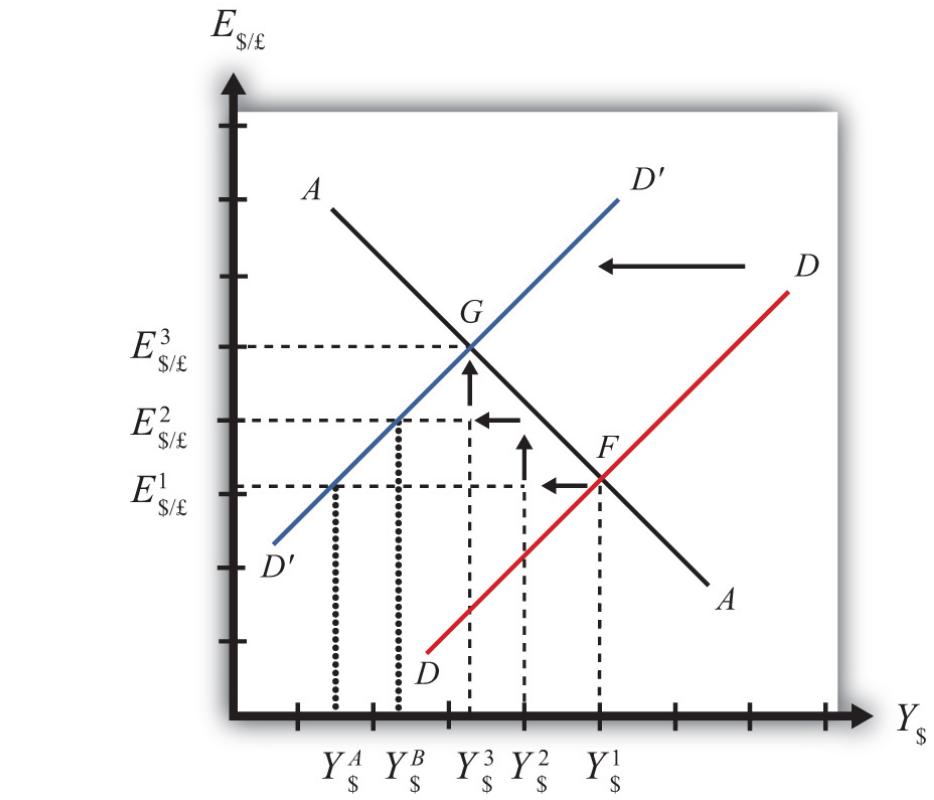
Although an equilibrium may never be reached in the real-world economy, the model remains very useful in understanding how changes in some variables will affect the behavior of agents and influence other variables. The model in essence offers us the opportunity to conduct experiments in simplified settings. Changing one exogenous variable and inferring its effect is a comparative statics *experiment* because of the *ceteris paribus* assumption. *Ceteris paribus* allows us to isolate one change and work through its impact with certainty that nothing else could influence the result.

Below, we'll consider adjustment to two changes: a reduction in investment demand, which shifts the DD curve, and an increase in foreign interest rates, which shifts the AA curve.

Reduction in Investment

Consider adjustment to a decrease in investment demand. Begin with an original superequilibrium, where DD crosses AA at point F with GNP at $Y_{\1 and exchange rate at $E_{\$/\text{£}}^1$. When investment decreases, ceteris paribus, the DD curve shifts to the left, as was shown in [Chapter 20 "The AA-DD Model", Section 20.3 "Shifting the DD Curve"](#). This shift is shown in [Figure 20.9 "Effects of an Investment Demand Decrease in the AA-DD Model"](#) as a shift from DD to D'D'.

Figure 20.9 Effects of an Investment Demand Decrease in the AA-DD Model



The quick result is that the equilibrium shifts to point G, GNP falls to $Y_{\3 , and the exchange rate rises to $E_{\$/\text{£}}^3$. The increase in the exchange rate represents a depreciation of the U.S. dollar value. However, this result does not explain the adjustment process, so let's take a more careful look at how the economy gets from point F to G.

Step 1: When investment demand falls, aggregate demand falls short of aggregate supply, leading to a buildup of inventories. Firms respond by cutting back supply, and GNP slowly begins to fall. Initially, there is no change in the exchange rate. On

the graph, this is represented by a leftward shift from the initial equilibrium at point F ($Y_{\1 to $Y_{\2). Adjustment to changes in aggregate demand will be gradual, perhaps taking several months or more to be fully implemented.

Step 2: As GNP falls, it causes a decrease in U.S. interest rates. With lower interest rates, the rate of return on U.S. assets falls below that in the United Kingdom and international investors shift funds abroad, leading to a dollar depreciation (pound appreciation)—that is, an increase in the exchange rate $E_{\$/\text{£}}$. This moves the economy upward, back to the AA curve. The adjustment in the asset market will occur quickly after the change in interest rates, so the leftward shift from point F in the diagram results in adjustment upward to regain equilibrium in the asset market on the AA curve.

Step 3: Continuing reductions in GNP caused by excess aggregate demand, results in continuing decreases in interest rates and rates of return, repeating the stepwise process above until the new equilibrium is reached at point G in the diagram.

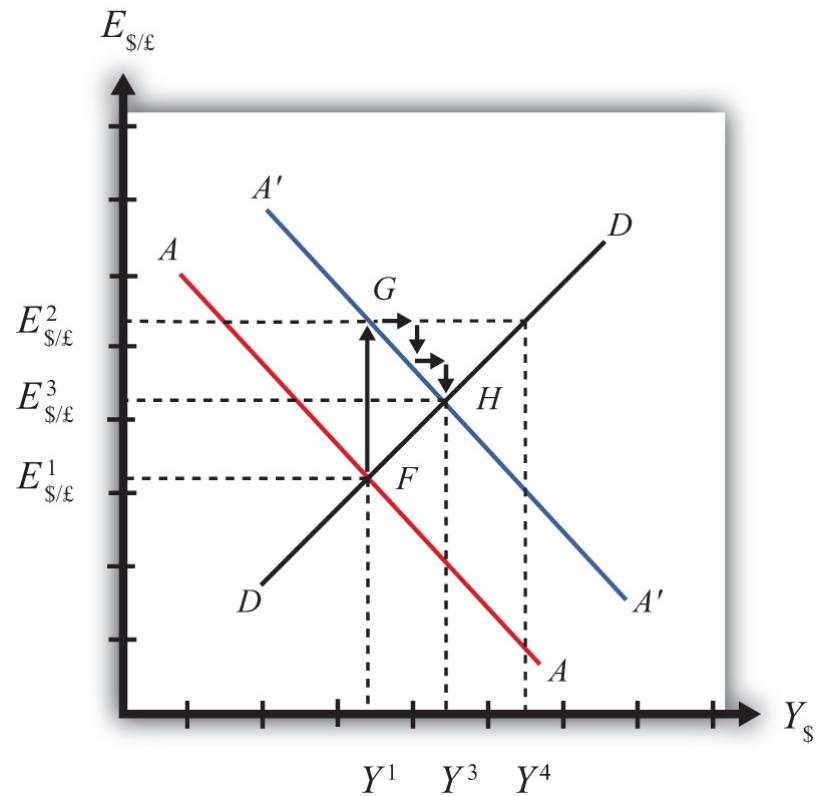
During the adjustment process, there are several other noteworthy changes taking place. At the initial equilibrium, when investment demand first falls, aggregate supply exceeds demand by the difference of $Y_{\$}^2 - Y_{\A . Adjustment in the goods market will be trying to reach equilibrium by getting back to the DD curve. However, the economy will never get to $Y_{\A . That's because the asset market will adjust in the meantime. As GNP falls, the exchange rate is pushed up to get back onto the AA curve. Remember, that asset market adjustment takes place quickly after an interest rate change (perhaps in several hours or days), while goods market adjustment can take months. When the exchange rate rises, the dollar depreciation makes foreign goods more expensive and reduces imports. It also makes U.S. goods cheaper to foreigners and stimulates exports, both of which cause an increase in current account demand. This change in demand is represented as a movement along the new $D'D'$ curve. Thus when the exchange rate rises up to $E_{\$/\text{£}}^2$ during the adjustment process, aggregate demand will have risen from $Y_{\A to $Y_{\B along the new $D'D'$ curve. In other words, the “target” for GNP adjustment moves closer as the exchange rate rises. In the end, the target for GNP reaches $Y_{\3 just as the exchange rate rises to $E_{\$/\text{£}}^3$.

Increase in Foreign Interest Rates

Consider adjustment to an increase in the foreign interest rate, i_E . Begin with an original superequilibrium where DD crosses AA at point F with GNP at Y^1 and exchange rate at $E_{\$/\text{£}}^1$. When the foreign interest rate increases, ceteris paribus, the AA curve shifts upward, as was shown in [Chapter 20 "The AA-DD Model", Section](#)

20.5 "Shifting the AA Curve". This shift is shown in Figure 20.10 "Effects of an Increase in Foreign Interest Rates in the AA-DD Model" as a shift from AA to A'A'.

Figure 20.10 Effects of an Increase in Foreign Interest Rates in the AA-DD Model



The quick result is that the equilibrium shifts to point H, GNP rises to Y^3 , and the exchange rate rises to E^3 . The increase in the exchange rate represents a depreciation of the U.S. dollar value.

The convenience of the graphical approach is that it allows us to quickly identify the final outcome using only our knowledge about the mechanics of the AA-DD diagram. However, this quick result does not explain the adjustment process, so let's take a more careful look at how the economy gets from point F to H.

Step 1: When the foreign interest rate (i_E) rises, the rate of return on foreign British assets rises above the rate of return on domestic U.S. assets in the foreign exchange market. This causes an immediate increase in the demand for foreign British currency, causing an appreciation of the pound and a depreciation of the U.S. dollar. Thus the exchange rate ($E_{\$/£}$) rises. This change is represented by the movement from point F to G on the AA-DD diagram. The AA curve shifts up to

reflect the new set of asset market equilibriums corresponding to the now-higher foreign interest rate. Since the foreign exchange market adjusts very swiftly to changes in interest rates, the economy will not remain off the new A'A' curve for very long.

Step 2: Now that the exchange rate has risen to $E_{\$/\text{£}}^2$, the real exchange has also increased. This implies foreign goods and services are relatively more expensive while U.S. G&S are relatively cheaper. This will raise demand for U.S. exports, curtail demand for U.S. imports, and result in an increase in current account and thereby aggregate demand. Note that the new equilibrium demand at exchange rate is temporarily at GNP level Y^4 , which is on the DD curve given the exchange rate $E_{\$/\text{£}}^2$. Because aggregate demand exceeds aggregate supply, inventories will begin to fall, stimulating an increase in production and thus GNP. This is represented by a rightward shift from point G (small arrow).

Step 3: As GNP rises, so does real money demand, causing an increase in U.S. interest rates. With higher interest rates, the rate of return on U.S. assets rises above that in the United Kingdom and international investors shift funds back to the United States, leading to a dollar appreciation (pound depreciation), or the decrease in the exchange rate ($E_{\$/\text{£}}$). This moves the economy downward, back to the A'A' curve. The adjustment in the asset market will occur quickly after the change in interest rates. Thus the rightward shift from point G in the diagram results in quick downward adjustment to regain equilibrium in the asset market on the A'A' curve, as shown.

Step 4: Continuing increases in GNP caused by excess aggregate demand, results in continuing increases in U.S. interest rates and rates of return, repeating the stepwise process above until the new equilibrium is reached at point H in the diagram.

During the adjustment process, there are several other noteworthy changes taking place. At point G, aggregate demand exceeds supply by the difference $Y^4 - Y^1$. Adjustment in the goods market will be trying to reach equilibrium by getting back to the DD curve. However, the economy will never get to Y^4 . That's because the asset market will adjust during the transition. As GNP rises, the exchange rate is gradually pushed down to get back onto the A'A' curve. When the exchange rate falls, the dollar appreciation makes foreign goods cheaper, raising imports. It also makes U.S. goods more expensive to foreigners, reducing exports—both of which cause a decrease in current account demand. This change in demand is represented as a movement along the DD curve. Thus when the exchange rate falls during the adjustment process, aggregate demand falls from Y^4 along the DD curve. In other

words, the “target” for GNP adjustment moves closer as the exchange rate falls. In the end, the target for GNP reaches Y^3 just as the exchange rate falls to E/£^3$.

KEY TAKEAWAYS

- Adjustment in the asset market occurs quickly, whereas adjustment in the G&S market occurs much more slowly.
- In the AA-DD model, a decrease in investment demand ultimately reduces GNP and raises the exchange rate, which, as defined, means a depreciation of the dollar.
- In the AA-DD model, an increase in foreign interest rates ultimately raises GNP and raises the exchange rate, which, as defined, means a depreciation of the dollar.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, the final effect on equilibrium GNP following an increase in investment demand in the AA-DD model.
 - b. Of *increase, decrease or stay the same*, the immediate effect on E/£$ following an increase in investment demand in the AA-DD model.
 - c. Of *increase, decrease, or stay the same*, the final effect on equilibrium GNP following a decrease in foreign interest rates in the AA-DD model.
 - d. Of *increase, decrease, or stay the same*, the immediate effect on E/£$ following a decrease in British interest rates in the AA-DD model.
 - e. Of *faster, slower, or the same rate*, this describes the speed of adjustment to a DD curve relative to an AA curve.

20.8 AA-DD and the Current Account Balance

LEARNING OBJECTIVE

- Derive a graphical mechanism in the AA-DD model to represent the effects of changes in the superequilibrium on the current account balance.

In later chapters we will use the AA-DD model to describe the effects of policy changes on macroeconomic variables in an open economy. The two most important macro variables are the exchange rate and the current account (trade) balance. The effects of changes on the exchange rate are vividly portrayed in the AA-DD diagram since this variable is plotted along the vertical axis and its value is determined as an element of the equilibrium. The current account (CA) variable is not displayed in the AA-DD diagram, but with some further thought we can devise a method to identify the current account balance at different positions in the AA-DD diagram.

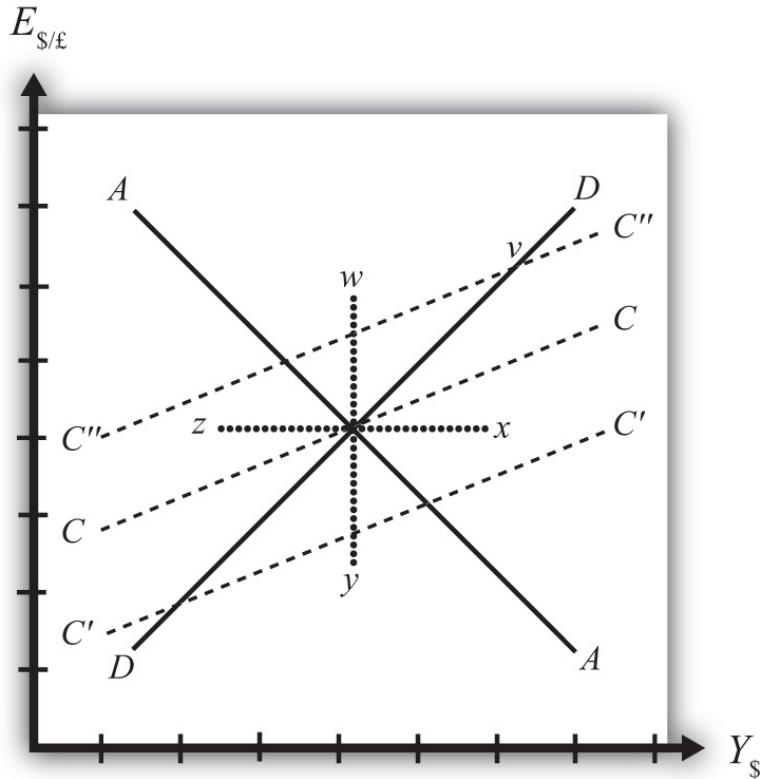
First, note that there is no “equilibrium” current account balance in a floating exchange rate system. Any balance on the current account is possible because any balance can correspond to balance on the balance of payments. The balance of payments is made up of two broad subaccounts: the current account and the financial account, the sum of whose balances must equal zero. When the balances sum to zero, the foreign demand for domestic goods, services, income, and assets is equal to domestic supply of goods, services, income, and assets. Thus there must always be “balance” on the balance of payments regardless of the balances on the individual subaccounts.

Iso-CAB Line

An **iso-CAB line**⁵ is a line drawn on an AA-DD diagram, [Figure 20.11 "Iso-CAB Lines in an AA-DD Diagram"](#), representing a set of points along which the current account balance (CAB) is the same. Note that “iso” is a prefix that means *the same*. In the adjoining diagram, we have superimposed three-dotted iso-CAB lines labeled *CC*, *C'C'*, and *C''C''*. Each line represents a set of GNP and exchange rate combinations that generate the same balance on the current account. The higher the CAB line, the larger is the balance on the current account. Thus the CAB balance on *C''C''* is greater than the balance along *CC*. Also note that each CAB line is positively sloped with a slope less than the slope of the DD curve. Next, we’ll continue with a justification for this description.

5. A line drawn on an AA-DD diagram representing a set of points along which the current account balance (CAB) is the same.

Figure 20.11 Iso-CAB Lines in an AA-DD Diagram



Justifying the Shape of the Iso-CAB Line

Consider the superequilibrium point at the intersection of AA and DD. The positions of these two curves are determined by the values of the exogenous variables in the model, including the domestic price level ($P_{\$}$), the foreign price level ($P_{\text{£}}$), tax revenues (T), and transfer payments (TR), among others. The intersection of the two curves determines the equilibrium GNP level ($Y_{\$}$) and the exchange rate ($E_{\$/\text{£}}$) (not labeled in diagram). Recall from [Chapter 20 "The AA-DD Model"](#), [Section 20.2 "Derivation of the DD Curve"](#) that the DD curve is derived from the aggregate demand function, one component of which is the current account function. The current account function, as shown below, is a function of all the variables listed immediately above:

$$CA^D \left(\frac{E_{\$/\text{£}} P_{\text{£}}}{P_{\$}}, Y - T + TR \right) = K.$$

Thus at the intersection of AA and DD there are presumed known values for the exogenous variables and determined values for the endogenous variables, $E_{\$/\text{£}}$ and $Y_{\$}$.

All these values could, in principle, be plugged into the current account demand function (CA^D) to determine the CA balance at the equilibrium. Let's assume that value is given by K , as shown in the above expression.

Now let's consider movements in the superequilibrium to other points on the diagram. Let's suppose that the equilibrium moved to point x directly to the right. That could arise from a rightward shift of DD and an upward shift of AA. We will also assume that this shift did not arise due to changes in $P_{\$}$, $P_{\text{£}}$, T , or TR , the other exogenous variables that affect the current account. (More on this issue below.) One possibility is an increase in the money supply and an increase in investment demand. Note that these shifts are not depicted.

At point x , GNP is higher while the exchange rate and the other exogenous variables are the same as before. Since an increase in $Y_{\$}$ raises disposable income, which reduces current account demand, the current account balance must be at a lower level at point x compared to the initial equilibrium.

If the equilibrium had shifted to point z instead, then GNP is lower while the exchange rate and the other exogenous variables are the same as before. Since a decrease in $Y_{\$}$ lowers disposable income, which raises current account demand, the current account balance must be at a higher level at point z compared to the initial equilibrium.

Next, suppose the equilibrium had shifted to point y instead. In this case, the exchange rate ($E_{\$/\text{£}}$) is lower while GNP and the other exogenous variables are the same as before. Since a decrease in $E_{\$/\text{£}}$ reduces the real exchange rate, which reduces current account demand, the current account balance must be at a lower level at point y compared to the initial equilibrium.

Finally, suppose the equilibrium had shifted to point w . In this case, the exchange rate, $E_{\$/\text{£}}$, is higher while GNP and the other exogenous variables are the same as before. Since an increase in $E_{\$/\text{£}}$ raises the real exchange rate, which increases current account demand, the current account balance must be at a higher level at point y compared to the initial equilibrium.

Since a movement to w and z results in an increase in the current account balance, while a shift to x or y causes a reduction in the balance, the line representing a constant CAB must be positively sloped.

Another way to see this is to use the CA^D function above. Suppose the CAB is originally at the value K . If the exchange rate ($E_{\$/£}$) rises, *ceteris paribus*, then CA will rise. We can now ask how GNP would have to change to get back to a CA balance of K . Clearly, if Y rises, disposable income rises and the current account balance falls. Raise GNP by precisely the right amount, and we can get the CAB back to K . Thus an increase in $E_{\$/£}$ must accompany an increase in GNP to maintain a fixed current account balance and therefore an iso-CAB line must be positively sloped.

The last thing we need to show is that the iso-CAB line is less steeply sloped than the DD line. Suppose the economy moved to a point such as v , which is on the same DD curve as the original superequilibrium. Recall from [Chapter 20 "The AA-DD Model"](#), [Section 20.2 "Derivation of the DD Curve"](#), the DD curve is derived from a change in the exchange rate and its effect on equilibrium GNP in the G&S market alone. The increase in the exchange rate causes an increase in current account demand through its effect on the real exchange rate. This causes an increase in aggregate demand, which inspires the increase in GNP. When equilibrium is reached in the G&S market, at point v , aggregate supply, Y , will equal aggregate demand and the following expression must hold:

$$Y_{\$} = C^D(Y_{\$} - \bar{T} + TR) + I_0 + G_0 + CA^D\left(\frac{E_{\$/£}^+ P_{£}}{P_{\$}}, Y_{\$} - \bar{T} + TR\right).$$

The left side is aggregate supply given by the equilibrium value at point v and the right side is aggregate demand. Since GNP is higher at v , consumption demand (C^D) must also be higher. However, because the marginal propensity to consume is less than one, not all the extra GNP will be spent on consumption goods; some will be saved. Nevertheless, aggregate demand (on the right side) must rise up to match the increase in supply on the left side. Since all the increase in demand cannot come from consumption, the remainder must come from the current account. This implies that a movement along the DD curve to v results in an increase in the current account balance. It also implies that the iso-CAB line must be less steeply sloped than the DD curve.

Using the Iso-CAB Line

The iso-CAB line can be used to assess the change in the country's current account balance from any exogenous variable change except changes in $P_{\$}$, $P_{£}$, T , and TR .

The reason we must exclude these variables is because the current account demand function is also dependent on these exogenous variables. If tax revenues increased, for example, all the iso-CAB lines would shift, making it much more difficult to pinpoint the final effect on the current account balance.

However, for monetary policy changes and government spending fiscal policy changes, the iso-CAB line will work. Anytime the superequilibrium shifts above the original iso-CAB line, the economy will move onto another iso-CAB line with a higher balance. (This is like the shift to point v in [Figure 20.11 "Iso-CAB Lines in an AA-DD Diagram"](#).) Recall that the $CA = EX - IM$, which can be positive or negative. If CAB were in surplus originally, an increase in the CAB (as with a movement to v) would imply an increase in the CA surplus. However, if the CAB were in deficit originally, then an increase in CAB implies a reduction in the deficit. If the increase in the CAB were sufficiently large, the CAB could move from deficit to surplus.

In a similar way, anytime the superequilibrium shifts below an initial iso-CAB line, the CAB surplus will fall, or the CAB deficit will rise.

Remember that the iso-CAB line is only used a reference to track the current account balance. The iso-CAB line is not used to determine the superequilibrium. For this reason, the iso-CAB line is plotted as a dashed line rather than a solid line.

KEY TAKEAWAYS

- An iso-CAB line is a line drawn on an AA-DD diagram, representing a set of points along which the current account balance (CAB) is the same.
- An iso-CAB line is positively sloped and with a slope that is less than the slope of the DD curve.
- The iso-CAB line can be used to assess the change in the country's current account balance from any exogenous variable change except changes in $P_{\$}$, P_E , T , and TR .

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *greater than, less than, or equal*, the current account balance for an exchange rate and GNP combination that lies above an iso-CAB line relative to a combination that lies on the line.
 - b. Of *greater than, less than, or equal*, the current account surplus for an exchange rate and GNP combination that lies below an iso-CAB line relative to the surplus for a combination that lies on the line.
 - c. Of *greater than, less than, or equal*, the current account deficit for an exchange rate and GNP combination that lies below an iso-CAB line relative to the deficit for a combination that lies on the line.
 - d. Of *higher, lower, or equal*, the position of an iso-CAB line for a country with a current account deficit relative to an iso-CAB line when the country runs a surplus.
 - e. Of *positive, negative, or zero*, this describes the slope of an iso-CAB line.
 - f. Of *steeper, flatter, or the same*, this describes an iso-CAB line relative to a DD curve.

Chapter 21

Policy Effects with Floating Exchange Rates

The effects of government policies on key macroeconomic variables are an important issue in international finance. The AA-DD model constructed in [Chapter 20 "The AA-DD Model"](#) is used in this chapter to analyze the effects of fiscal and monetary policy under a regime of floating exchange rates. The results are more comprehensive than the previous analyses of the same policies because they take into account all the between-market effects across the money market, the foreign exchange (Forex) market, and the goods and services (G&S) market.

21.1 Overview of Policy with Floating Exchange Rates

LEARNING OBJECTIVE

1. Preview the comparative statics results from the AA-DD model with floating exchange rates.

This chapter uses the AA-DD model to describe the effects of fiscal and monetary policy under a system of floating exchange rates. Fiscal and monetary policies are the primary tools governments use to guide the macroeconomy. In introductory macroeconomics courses, students learn how fiscal and monetary policy levers can be used to influence the level of gross national product (GNP), the inflation rate, the unemployment rate, and interest rates. In this chapter, that analysis is expanded to an open economy (i.e., one open to trade) and to the effects on exchange rates and current account balances.

Results

Using the AA-DD model, several important relationships between key economic variables are shown:

1. An increase in the money supply in a country.
2. An exchange rate system in which the value of a country's currency is determined by the supply and demand for one currency in exchange for another in a private market operated by major international banks.
3. A decrease in the money supply in a country.
4. An increase in government spending or transfer payments and/or a decrease in tax revenues.
5. A decrease in government spending or transfer payments and/or an increase in tax revenues.

- **Expansionary monetary policy**¹ ($\uparrow M^S$) causes an increase in GNP and a depreciation of the domestic currency in a **floating exchange rate system**² in the short run.
- **Contractionary monetary policy**³ ($\downarrow M^S$) causes a decrease in GNP and an appreciation of the domestic currency in a floating exchange rate system in the short run.
- **Expansionary fiscal policy**⁴ ($\uparrow G, \uparrow TR$, or $\downarrow T$) causes an increase in GNP and an appreciation of the domestic currency in a floating exchange rate system.
- **Contractionary fiscal policy**⁵ ($\downarrow G, \downarrow TR$, or $\uparrow T$) causes a decrease in GNP and a depreciation of the domestic currency in a floating exchange rate system.
- In the long run, once inflation effects are included, expansionary monetary policy ($\uparrow M^S$) in a full employment economy causes no long-term change in GNP and a depreciation of the domestic currency in a floating exchange rate system. In the transition, the exchange rate overshoots its long-run target and GNP rises then falls.

- A **sterilized foreign exchange intervention**⁶ will have no effect on GNP or the exchange rate in the AA-DD model, unless international investors adjust their expected future exchange rate in response.
- A central bank can influence the exchange rate with direct Forex interventions (buying or selling domestic currency in exchange for foreign currency). To sell foreign currency and buy domestic currency, the central bank must have a stockpile of foreign currency reserves.
- A central bank can also influence the exchange rate with indirect open market operations (buying or selling domestic treasury bonds). These transactions work through money supply changes and their effect on interest rates.
- Purchases (sales) of foreign currency on the Forex will raise (lower) the domestic money supply and cause a secondary indirect effect upon the exchange rate.

Connections

The AA-DD model was developed to describe the interrelationships of macroeconomic variables within an open economy. Since some of these macroeconomic variables are controlled by the government, we can use the model to understand the likely effects of government policy changes. The two main levers the government controls are monetary policy (changes in the money supply) and fiscal policy (changes in the government budget). In this chapter, the AA-DD model is applied to understand government policy effects in the context of a floating exchange rate system. In [Chapter 23 "Policy Effects with Fixed Exchange Rates"](#), we'll revisit these same government policies in the context of a fixed exchange rate system.

It is important to recognize that these results are what “would” happen under the full set of assumptions that describe the AA-DD model. These effects may or may not happen in reality. Despite this problem, the model surely captures some of the simple cause-and-effect relationships and therefore helps us to understand the broader implications of policy changes. Thus even if in reality many more elements not described in the model may act to influence the key endogenous variables, the AA-DD model at least gives a more complete picture of some of the expected tendencies.

6. Occurs when a central bank counters direct intervention in the Forex with a simultaneous offsetting transaction in the domestic bond market.

KEY TAKEAWAYS

- The main objective of the AA-DD model is to assess the effects of monetary and fiscal policy changes.
- It is important to recognize that these results are what “would” happen under the full set of assumptions that describes the AA-DD model; they may or may not accurately describe actual outcomes in actual economies.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP in the short run if government spending decreases in the AA-DD model with floating exchange rates.
 - b. Of *increase, decrease, or stay the same*, this is the effect on the domestic currency value in the short run if government spending decreases in the AA-DD model with floating exchange rates.
 - c. Of *increase, decrease, or stay the same*, this is the effect on the foreign currency value (vis-à-vis the domestic) in the short run if domestic government spending decreases in the AA-DD model with floating exchange rates.
 - d. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP in the short run if the nominal money supply decreases in the AA-DD model with floating exchange rates.
 - e. Of *increase, decrease, or stay the same*, this is the effect on the domestic currency value in the short run if the nominal money supply decreases in the AA-DD model with floating exchange rates.
 - f. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP in the long run if the nominal money supply increases in the AA-DD model with floating exchange rates.
 - g. Of *increase, decrease, or stay the same*, this is the effect on the domestic currency value in the long run if the nominal money supply increases in the AA-DD model with floating exchange rates.

21.2 Monetary Policy with Floating Exchange Rates

LEARNING OBJECTIVES

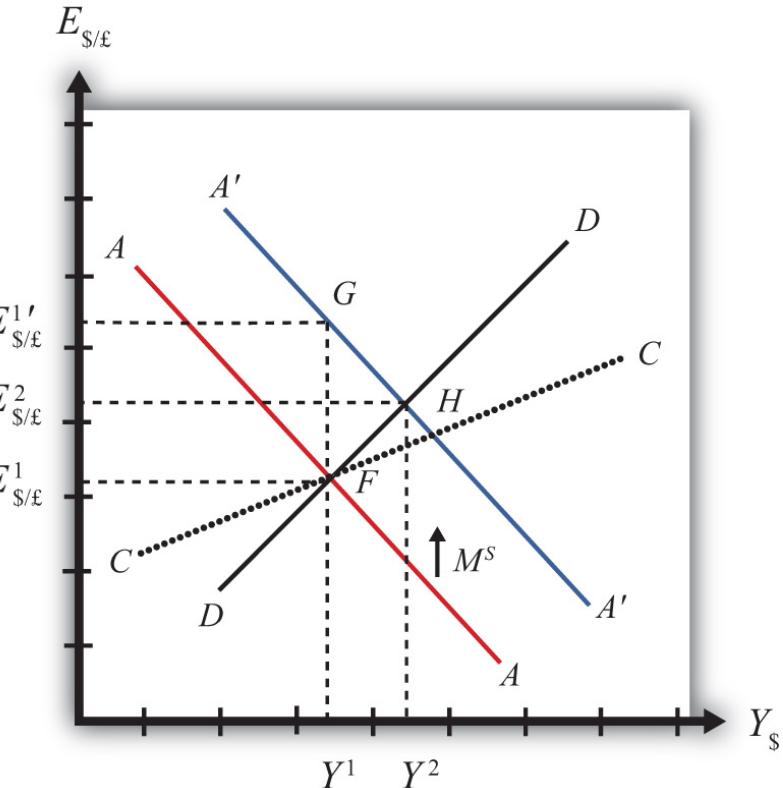
1. Learn how changes in monetary policy affect GNP, the value of the exchange rate, and the current account balance in a floating exchange rate system in the context of the AA-DD model.
2. Understand the adjustment process in the money market, the foreign exchange market, and the G&S market.

In this section, we use the AA-DD model to assess the effects of monetary policy in a floating exchange rate system. Recall from [Chapter 18 "Interest Rate Determination"](#) that the money supply is effectively controlled by a country's central bank. In the case of the United States, this is the Federal Reserve Board, or the Fed for short. When the money supply increases due to action taken by the central bank, we refer to it as expansionary monetary policy. If the central bank acts to reduce the money supply, it is referred to as contractionary monetary policy. Methods that can be used to change the money supply are discussed in [Chapter 18 "Interest Rate Determination"](#), [Section 18.5 "Controlling the Money Supply"](#).

Expansionary Monetary Policy

Suppose the economy is originally at a superequilibrium shown as point F in [Figure 21.1 "Expansionary Monetary Policy in the AA-DD Model with Floating Exchange Rates"](#). The original GNP level is Y^1 and the exchange rate is $E_{\$/£}^1$. Next, suppose the U.S. central bank (or the Fed) decides to expand the money supply. As shown in [Chapter 20 "The AA-DD Model"](#), [Section 20.5 "Shifting the AA Curve"](#), money supply changes cause a shift in the AA curve. More specifically, an increase in the money supply will cause AA to shift upward (i.e., $\uparrow M^S$ is an AA up-shifter). This is depicted in the diagram as a shift from the red AA to the blue A'A' line.

Figure 21.1 Expansionary Monetary Policy in the AA-DD Model with Floating Exchange Rates



There are several different levels of detail that can be provided to describe the effects of this policy. Below, we present three descriptions with increasing degrees of completeness. First the quick result, then the quick result with the transition process described, and finally the complete adjustment story.

Quick Result

The increase in AA causes a shift in the superequilibrium point from F to H . In adjusting to the new equilibrium at H , GNP rises from Y^1 to Y^2 and the exchange rate increases from $E_{\$/\£}^1$ to $E_{\$/\£}^2$. The increase in the exchange rate represents an increase in the British pound value and a decrease in the U.S. dollar value. In other words, it is an appreciation of the pound and a depreciation of the dollar. Since the final equilibrium point H is above the initial iso-CAB line CC , the current account balance increases. (See [Chapter 20 "The AA-DD Model"](#), [Section 20.8 "AA-DD and the Current Account Balance"](#) for a description of CC .) If the CAB were in surplus at F , then the surplus increases; if the CAB were in deficit, then the deficit falls. Thus U.S. expansionary monetary policy causes an increase in GNP, a depreciation of the U.S. dollar, and an increase in the current account balance in a floating exchange rate system according to the AA-DD model.

Transition Description

Consider the upward shift of the AA curve due to the increase in the money supply. Since exchange rates adjust much more rapidly than GNP, the economy will initially adjust back to the new A'A' curve before any change in GNP occurs. That means the first adjustment will be from point F to point G directly above. The exchange rate will increase from $E_{\$/\text{£}}^1$ to $E_{\$/\text{£}}^{1'}$, representing a depreciation of the U.S. dollar.

Now at point G, the economy lies to the left of the DD curve. Thus GNP will begin to rise to get back to goods and services (G&S) market equilibrium on the DD curve. However, as GNP rises, the economy moves to the right above the A'A' curve, which forces a downward readjustment of the exchange rate to get back to A'A'. In the end, the economy will adjust in a stepwise fashion from point G to point H, with each rightward movement in GNP followed by a quick reduction in the exchange rate to remain on the A'A' curve. This process will continue until the economy reaches the superequilibrium at point H.

Notice that in the transition the exchange rate first rises to $E_{\$/\text{£}}^{1'}$. Above the rate it will ultimately reach $E_{\$/\text{£}}^2$ before settling back to superequilibrium value. This is an example of exchange rate overshooting. In the transition, the exchange rate overshoots its ultimate long-run value.

Exchange rate overshooting⁷ is used as one explanation for the volatility of exchange rates in floating markets. If many small changes occur frequently in an economy, the economy may always be in transition moving to a superequilibrium. Because of the more rapid adjustment of exchange rates, it is possible that many episodes of overshooting—both upward and downward—can occur in a relatively short period.

Complete Adjustment Story

Step 1: When the money supply increases, real money supply will exceed real money demand in the economy. Since households and businesses hold more money than they would like, at current interest rates, they begin to convert liquid money assets into less-liquid nonmoney assets. This raises the supply of long-term deposits and the amount of funds available for banks to loan. More money to lend will lower average U.S. interest rates, which in turn will result in a lower U.S. rate of return in the Forex market. Since $\text{RoR}_{\$} < \text{RoR}_{\text{£}}$ now, there will be an immediate increase in the demand for foreign British currency, thus causing an appreciation of the pound and a depreciation of the U.S. dollar. Thus the exchange rate ($E_{\$/\text{£}}$) rises. This change is represented by the movement from point F to G on the AA-DD diagram. The AA

7. Occurs when an exchange rate, in the process of adjusting to a new equilibrium, either rises above the final equilibrium value before falling back again, or falls below the final equilibrium value before rising up again.

curve has shifted up to reflect the new set of asset market equilibria corresponding to the higher U.S. money supply. Since the money market and foreign exchange (Forex) markets adjust very swiftly to the money supply change, the economy will not remain off the new A'A' curve for very long.

Step 2: Now that the exchange rate has risen to $E_{\$/\text{£}}^{1'}$, the real exchange has also increased. This implies foreign goods and services are relatively more expensive while U.S. G&S are relatively cheaper. This will raise demand for U.S. exports, curtail demand for U.S. imports, and result in an increase in current account and, thereby, aggregate demand. Because aggregate demand exceeds aggregate supply, inventories will begin to fall, stimulating an increase in production and thus GNP. This is represented by a rightward shift from point G.

Step 3: As GNP rises, so does real money demand, causing an increase in U.S. interest rates. With higher interest rates, the rate of return on U.S. assets rises above that in the United Kingdom, and international investors shift funds back to the United States, resulting in a dollar appreciation (pound depreciation)—that is, a decrease in the exchange rate ($E_{\$/\text{£}}$). This moves the economy downward, back to the A'A' curve. The adjustment in the asset market will occur quickly after the change in interest rates. Thus the rightward shift from point G in the diagram results in quick downward adjustment to regain equilibrium in the asset market on the A'A' curve, as shown in the figure.

Step 4: Continuing increases in GNP caused by excess aggregate demand, results in continuing increases in U.S. interest rates and rates of return, repeating the stepwise process above until the new equilibrium is reached at point H in the diagram.

Step 5: The equilibrium at H lies to the northeast of F along the original DD curve. As shown in [Chapter 20 "The AA-DD Model"](#), [Section 20.8 "AA-DD and the Current Account Balance"](#), the equilibrium at H lies above the original iso-CAB line. Therefore, the current account balance will rise.

Contractionary Monetary Policy

Contractionary monetary policy corresponds to a decrease in the money supply. In the AA-DD model, a decrease in the money supply shifts the AA curve downward. The effects will be the opposite of those described above for expansionary monetary policy. A complete description is left for the reader as an exercise.

The quick effects, however, are as follows. U.S. contractionary monetary policy will cause a reduction in GNP and a reduction in the exchange rate, $E_{\$/\text{£}}$, implying an appreciation of the U.S. dollar and a decrease in the current account balance.

KEY TAKEAWAYS

- The U.S. expansionary monetary policy causes an increase in GNP, a depreciation of the U.S. dollar, and an increase in the current account balance in a floating exchange rate system according to the AA-DD model.
- Contractionary monetary policy will cause a reduction in GNP and a reduction in the exchange rate ($E_{\$/\text{£}}$), implying an appreciation of the U.S. dollar and a decrease in the current account balance.

EXERCISES

1. Use the AA-DD model (not necessarily the diagram) to explain the *sequential* short-run adjustment process of an increase in the money supply on the following economic variables under floating exchange rates. (In other words, first answer how the money supply increase immediately affects the interest rate. Next, answer how the previous economic variable—i.e., the interest rate—affects the nominal exchange rate. Continue this process through investment.)
 - a. The interest rate
 - b. The nominal exchange rate
 - c. The real exchange rate
 - d. The current account balance
 - e. GNP
 - f. Disposable income
 - g. Consumption
 - h. Saving
 - i. Investment
2. Repeat the exercise above assuming a decrease in the money supply.
3. Suppose a country with floating exchange rates has a current account deficit that its government considers too large. Use an AA-DD diagram to show how monetary policy could be used to reduce the current account deficit. Does this action help or hinder its goal of maintaining low unemployment? Explain.

21.3 Fiscal Policy with Floating Exchange Rates

LEARNING OBJECTIVES

1. Learn how changes in fiscal policy affect GNP, the value of the exchange rate, and the current account balance in a floating exchange rate system in the context of the AA-DD model.
2. Understand the adjustment process in the money market, the Forex market, and the G&S market.

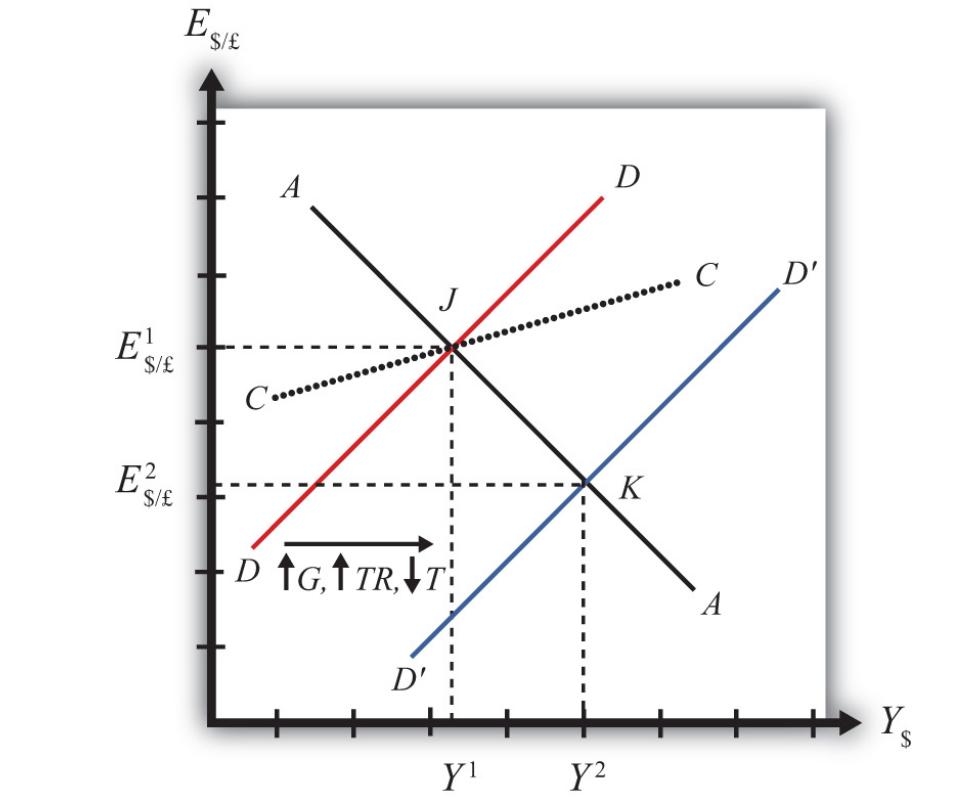
In this section, we use the AA-DD model to assess the effects of fiscal policy in a floating exchange rate system. Recall that fiscal policy refers to any change in expenditures or revenues within any branch of the government. This means any change in government spending—transfer payments, or taxes, by either federal, state, or local governments—represents a fiscal policy change. Since changes in expenditures or revenues will often affect a government budget balance, we can also say that a change in the government surplus or deficit represents a change in fiscal policy.

When government spending or transfer payments increase, or tax revenues decrease, we refer to it as expansionary fiscal policy. These actions would also be associated with an increase in the government budget deficit or a decrease in its budget surplus. If the government acts to reduce government spending or transfer payments, or increase tax revenues, it is referred to as contractionary fiscal policy. These actions would also be associated with a decrease in the government budget deficit, or an increase in its budget surplus.

Expansionary Fiscal Policy

Suppose the economy is originally at a superequilibrium shown as point J in [Figure 21.2 "Expansionary Fiscal Policy in the AA-DD Model with Floating Exchange Rates"](#). The original gross national product (GNP) level is Y^1 and the exchange rate is E/£^1$. Next, suppose the government decides to increase government spending (or increase transfer payments or decrease taxes). As shown in [Chapter 20 "The AA-DD Model"](#), [Section 20.3 "Shifting the DD Curve"](#), fiscal policy changes cause a shift in the DD curve. More specifically, an increase in government spending (or an increase in transfer payments or a decrease in taxes) will cause DD to shift rightward (i.e., $\uparrow G$, $\uparrow TR$, and $\downarrow T$ all are DD right-shifters). This is depicted in the diagram as a shift from the red DD to the blue $D'D'$ line.

Figure 21.2 Expansionary Fiscal Policy in the AA-DD Model with Floating Exchange Rates



There are several different levels of detail that can be provided to describe the effects of this policy. Below, we present three descriptions with increasing degrees of completeness: first the quick result, then the quick result with the transition process described, and finally the complete adjustment story.

Quick Result

The increase in DD causes a shift in the superequilibrium point from J to K. In adjusting to the new equilibrium at K, GNP rises from Y^1 to Y^2 and the exchange rate decreases from $E_{\$/£}^1$ to $E_{\$/£}^2$. The decrease in the exchange represents a decrease in the British pound value and an increase in the U.S. dollar value. In other words, it is a depreciation of the pound and an appreciation of the dollar. Since the final equilibrium point K is below the initial iso-CAB line CC, the current account balance decreases. (Caveat: this will be true for all fiscal expansions, but the iso-CAB line can only be used with an increase in G ; see [Chapter 20 "The AA-DD Model", Section 20.8 "AA-DD and the Current Account Balance"](#) for an explanation.) If the CAB were in surplus at J, then the surplus decreases; if the CAB were in deficit, then the deficit rises. Thus the U.S. expansionary fiscal policy causes an increase in the

U.S. GNP, an appreciation of the U.S. dollar, and a decrease in the current account balance in a floating exchange rate system according to the AA-DD model.

Transition Description

If the expansionary fiscal policy occurs because of an increase in government spending, then government demand for goods and services (G&S) will increase. If the expansionary fiscal policy occurs due to an increase in transfer payments or a decrease in taxes, then disposable income will increase, leading to an increase in consumption demand. In either case aggregate demand increases, and this causes the rightward shift in the DD curve. Immediately after aggregate demand increases, but before any adjustment has occurred at point J, the economy lies to the left of the new D'D' curve. Thus GNP will begin to rise to get back to G&S market equilibrium on the D'D' curve. However, as GNP rises, the economy will move above the AA curve, forcing a downward readjustment of the exchange rate to get back to asset market equilibrium on the AA curve. In the end, the economy will adjust in a stepwise fashion from point J to point K, with each rightward movement in GNP followed by a quick reduction in the exchange rate to remain on the AA curve. This process will continue until the economy reaches the superequilibrium at point K.

Complete Adjustment Story

Step 1: If the expansionary fiscal policy occurs because of an increase in government spending, then government demand for G&S will increase. If the expansionary fiscal policy occurs due to an increase in transfer payments or a decrease in taxes, then disposable income will increase, leading to an increase in consumption demand. In either case aggregate demand increases. Before any adjustment occurs, the increase in aggregate demand implies aggregate demand exceeds aggregate supply, which will lead to a decline in inventories. To prevent this decline, retailers (or government suppliers) will signal firms to produce more. As supply increases so does the GNP, and the economy moves to the right of point J.

Step 2: As GNP rises, so does real money demand, causing an increase in U.S. interest rates. With higher interest rates, the rate of return on U.S. assets rises above that in the United Kingdom and international investors shift funds back to the United States, resulting in a dollar appreciation (pound depreciation)—that is, a decrease in the exchange rate $E_{\$/£}$. This moves the economy downward, back to the AA curve. The adjustment in the asset market will occur quickly after the change in interest rates. Thus the rightward shift from point J in the diagram results in quick downward adjustment to regain equilibrium in the asset market on the AA curve, as shown.

Step 3: Continuing increases in GNP caused by excess aggregate demand, results in continuing increases in U.S. interest rates and rates of return, repeating the stepwise process above until the new equilibrium is reached at point K in the diagram.

Step 4: The equilibrium at K lies to the southeast of J along the original AA curve. As shown in [Chapter 20 "The AA-DD Model"](#), [Section 20.8 "AA-DD and the Current Account Balance"](#), the current account balance must be lower at K since both an increase in GNP and a dollar appreciation cause decreases in current account demand. Thus the equilibrium at K lies below the original iso-CAB line. However, this is only assured if the fiscal expansion occurred due to an increase in G.

If transfer payments increased or taxes were reduced, these would both increase disposable income and lead to a further decline in the current account balance. Thus also with these types of fiscal expansions, the current account balance is reduced; however, one cannot use the iso-CAB line to show it.

Contractionary Fiscal Policy

Contractionary fiscal policy corresponds to a decrease in government spending, a decrease in transfer payments, or an increase in taxes. It would also be represented by a decrease in the government budget deficit or an increase in the budget surplus. In the AA-DD model, a contractionary fiscal policy shifts the DD curve leftward. The effects will be the opposite of those described above for expansionary fiscal policy. A complete description is left for the reader as an exercise.

The quick effects, however, are as follows. U.S. contractionary fiscal policy will cause a reduction in GNP and an increase in the exchange rate ($E_{\$/£}$), implying a depreciation of the U.S. dollar.

KEY TAKEAWAYS

- Expansionary fiscal policy causes an increase in GNP, an appreciation of the currency, and a decrease in the current account balance in a floating exchange rate system according to the AA-DD model.
- Contractionary fiscal policy will cause a reduction in GNP, a depreciation of the currency, and an increase in the current account balance in a floating exchange rate system according to the AA-DD model.

EXERCISES

1. Suppose a country with floating exchange rates has a current account deficit that its government considers too large. Use an AA-DD diagram to show how fiscal policy could be used to reduce the current account deficit. Does this action help or hinder its goal of maintaining low unemployment?

2. The United States maintains a floating exchange rate. In the past few years, its government budget deficit has risen to a very high level. At the same time, its trade deficit has also become much larger.
 - a. Suppose the government reduces government spending to reduce the budget deficit. Assume the U.S. economy can be described with the AA-DD model. In the adjustment to the new equilibrium, the following variables will be affected in the order listed. Indicate whether each variable rises (+) or falls (-) during the adjustment process.

	Indicate + or -
Government Demand (G)	
Aggregate Demand (AD)	
Aggregate Supply ($Y_{\$}$)	
Real Money Demand ($L[i_{\$}, Y_{\$}]$)	
U.S. Interest Rates ($i_{\$}$)	
U.S. Rate of Return ($RoR_{\$}$)	
Exchange Rate ($E_{\$}/\underline{E}$)	
Foreign Rate of Return (RoR_E)	
Real Exchange Rate ($q_{\$}/\underline{E}$)	
Current Account Demand (CA^D)	
Aggregate Demand (AD)	

- b. Once the final short-run equilibrium is reached, indicate the effect of the decrease in government spending on the following variables:

	Indicate + or -
U.S. Government Budget Deficit	
U.S. Dollar Value	
U.S. Current Account Deficit	
U.S. GNP	

3. Consider the following actions/occurrences listed in the first column. For each one, use the AA-DD model to determine the impact on the variables from the twin-deficit identity listed along the top row. Consider the final equilibrium short-run effects. Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	Impact on			
	<i>S_p</i>	<i>I</i>	<i>IM - EX</i>	<i>G + TR - T</i>
a. A decrease in investment demand with floating ERs				
b. A decrease in investment demand floating ERs				
c. An increase in foreign interest rates under floating ERs				

	Impact on			
	S_p	I	$IM - EX$	$G + TR - T$
d. An increase in government demand under floating ERs				

21.4 Expansionary Monetary Policy with Floating Exchange Rates in the Long Run

LEARNING OBJECTIVES

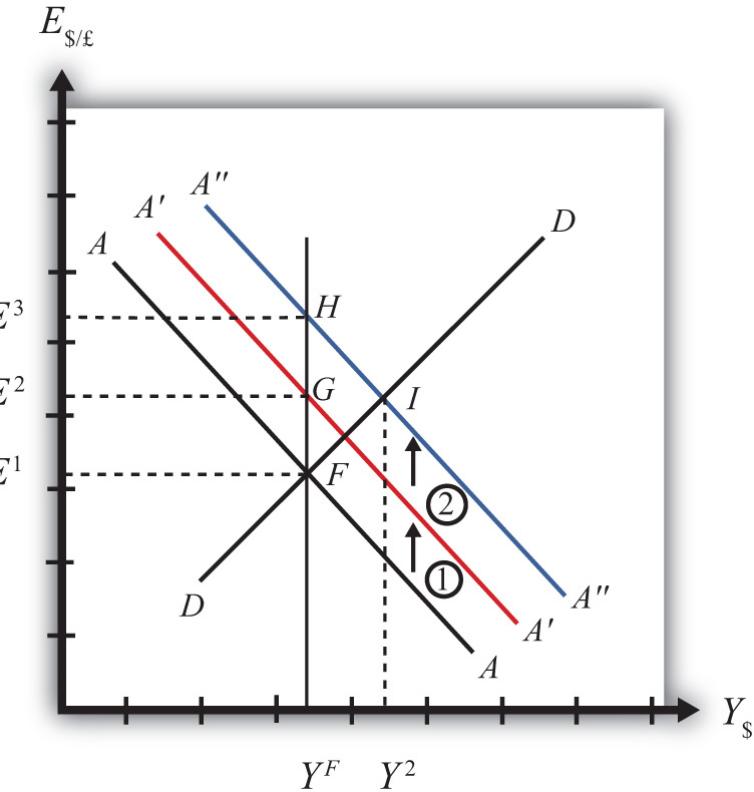
1. Learn how changes in monetary policy affect GNP and the value of the exchange rate in a floating exchange rate system in the context of the AA-DD model in the long run.
2. Understand the adjustment process in the money market, the Forex market, and the G&S market.

If expansionary monetary policy occurs when the economy is operating at full employment output, then the money supply increase will eventually put upward pressure on prices. Thus we say that eventually, or in the long run, the aggregate price level will rise and the economy will experience an episode of inflation in the transition. See [Chapter 18 "Interest Rate Determination"](#), [Section 18.14 "Money Supply and Long-Run Prices"](#) for a complete description of this process.

Here, we will describe the long-run effects of an increase in the money supply using the AA-DD model. We break up the effects into short-run and long-run components. In the short run, the initial money supply effects are felt and investor anticipations about future effects are implemented. In the long run, we allow the price level to rise.

Suppose the economy is originally at a superequilibrium, shown as point F in [Figure 21.3 "Expansionary Monetary Policy in the Long Run"](#). The original GNP level is Y^F , and the exchange rate is E^1 . Y^F represents the full-employment level of output, which also implies that the natural rate of unemployment prevails. Any movement of the economy to the right of Y^F will cause an eventual increase in the aggregate price level. Any movement to the left of Y^F causes an eventual decrease in the price level.

Figure 21.3 Expansionary Monetary Policy in the Long Run



Next, suppose the U.S. central bank (or the Fed) decides to expand the money supply. As shown in [Chapter 20 "The AA-DD Model"](#), [Section 20.5 "Shifting the AA Curve"](#), money supply changes cause a shift in the AA curve. More specifically, an increase in the money supply will cause AA to shift upward (i.e., $\uparrow M^S$ is an AA up-shifter). This is depicted in the diagram as a shift from the AA line to the red $A'A'$ line.

In the long-run adjustment story, several different changes in exogenous variables will occur sequentially, thus it is difficult to describe the quick final result, so we will only describe the transition process in partial detail.

Partial Detail

The increase in the money supply causes the first upward shift of the AA curve, shown as step 1 in the diagram. Since exchange rates adjust much more rapidly than gross national product (GNP), the economy will quickly adjust to the new $A'A'$ curve before any change in GNP occurs. That means the first adjustment will be

from point F to point G directly above. The exchange rate will increase from E^1 to E^2 , representing a depreciation of the U.S. dollar.

The second effect is caused by changes in investor expectations. Investors generally track important changes in the economy, including money supply changes, because these changes can have important implications for the returns on their investments. Investors who see an increase in money supply in an economy at full employment are likely to expect inflation to occur in the future. When investors expect future U.S. inflation, and when they consider both domestic and foreign investments, they will respond today with an increase in their expected future exchange rate ($E_{\$/\text{£}}^e$). There are two reasons to expect this immediate effect:

1. Investors are very likely to understand the story we are in the process of explaining now. As we will see below, the long-run effect of a money supply increase for an economy (initially, at full employment) is an increase in the exchange rate ($E_{\$/\text{£}}$)—that is, a depreciation of the dollar. If investors believe the exchange rate will be higher next year due to today's action by the Fed, then it makes sense for them to raise their expected future exchange rate in anticipation of that effect. Thus the average $E_{\$/\text{£}}^e$ will rise among investors who participate in the foreign exchange (Forex) markets.
2. Investors may look to the purchasing power parity (PPP) theory for guidance. PPP is generally interpreted as a long-run theory of exchange rate trends. If PPP holds in the long run, then $E_{\$/\text{£}} = P_{\$}/P_{\text{£}}$. In other words, the exchange rate will equal the ratio of the two countries' price levels. If $P_{\$}$ is expected to rise due to inflation, then PPP predicts that the exchange rate ($E_{\$/\text{£}}$) will also rise and the dollar will depreciate.

The timing of the change in $E_{\$/\text{£}}^e$ will depend on how quickly investors recognize the money supply change, compute its likely effect, and incorporate it into their investment plans. Since investors are typically very quick to adapt to market changes, the expectations effect should take place in the short run, perhaps long before the inflation ever occurs. In some cases, the expectations change may even occur before the Fed increases the money supply, if investors anticipate the Fed's action.

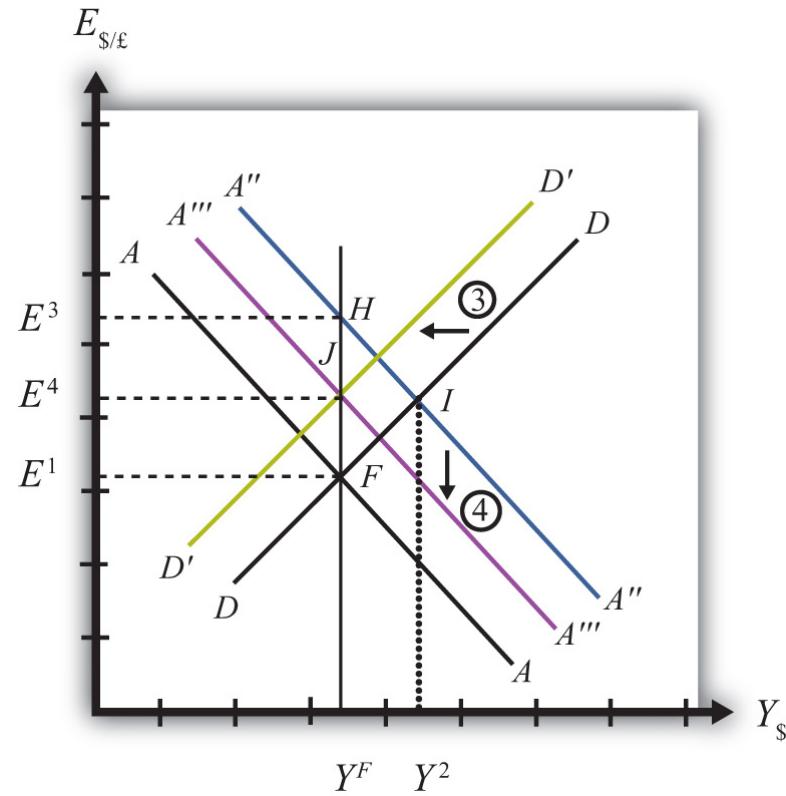
The increase in the expected exchange rate (this means a decrease in the expected future dollar value) causes a second upward shift of the AA curve, shown as step 2 in the diagram. Again, rapid exchange rate adjustment implies the economy will quickly adjust to the new A''A'' curve at point H directly above. The exchange rate

will now increase from E^2 to E^3 , representing a further depreciation of the U.S. dollar.

Once at point H , aggregate demand, which is on the DD curve to the right of H , exceeds aggregate supply, which is still at Y^F . Thus GNP will begin to rise to get back to G&S market equilibrium on the DD curve. However, as GNP rises, the economy moves above the $A''A''$ curve that forces a downward readjustment of the exchange rate to get back to asset market equilibrium on $A''A''$. In the end, the economy will adjust in a stepwise fashion from point H to point I , with each rightward movement in GNP followed by a quick reduction in the exchange rate to remain on the $A''A''$ curve. This process will continue until the economy reaches the temporary superequilibrium at point I .

The next effect occurs because GNP, now at Y^2 at point I , has risen above the full employment level at Y^F . This causes an increase in U.S. prices, meaning that $P_{\$}$ (the U.S. price level) begins to rise. The increase in U.S. prices has two effects as shown in [Figure 21.4 "Expansionary Monetary Policy in the Long Run, Continued"](#). An increase in $P_{\$}$ is both a DD left-shifter and an AA down-shifter.

Figure 21.4 Expansionary Monetary Policy in the Long Run, Continued



In step 3, we depict a leftward shift of DD to $D'D'$. DD shifts left because higher U.S. prices will reduce the real exchange rate. This makes U.S. G&S relatively more expensive compared with foreign G&S, thus reducing export demand, increasing import demand, and thereby reducing aggregate demand.

In step 4, we depict a downward shift of $A''A''$ to $A'''A'''$. AA shifts down because a higher U.S. price level reduces real money supply. As the real money supply falls, U.S. interest rates rise, leading to an increase in the rate of return for U.S. assets as considered by international investors. This in turn raises the demand for U.S. dollars on the Forex, leading to a dollar appreciation. Since this effect occurs for any GNP level, the entire AA curve shifts downward.

Steps 3 and 4 will both occur simultaneously, and since both are affected by the increase in the price level, it is impossible to know which curve will shift faster or precisely how far each curve will shift. However, we do know two things. First, the AA and DD shifting will continue as long as GNP remains above the full employment level. Once GNP falls to Y_F , there is no longer upward pressure on the price level and the shifting will cease. Second, the final equilibrium exchange rate must lie above the original exchange rate. This occurs because output will revert back to its original level, the price level will be higher, and according to PPP, eventually the exchange rate will have to be higher as well.

The final equilibrium will be at a point like J, which lies to the left of I. In this transition, the exchange rate will occasionally rise when DD shifts left and will occasionally fall when AA shifts down. Thus the economy will wiggle its way up and down, from point I to J. Once at point J, there is no reason for prices to rise further and no reason for a change in investor expectations. The economy will have reached its long-run equilibrium.

Note that one cannot use the iso-CAB line to assess the long-run effect on the current account balance. In the final adjustment, although the final equilibrium lies above the original iso-CAB line, in the long run the $P_{\$}$ changes will raise the iso-CAB lines, making it impossible to use these to identify the final effect.

However, in adjusting to the long-run equilibrium, the only two variables affecting the current account that will ultimately change are the exchange rate and the price level. If these two rise proportionally to each other, as they would if purchasing power parity held, then there will be no long-run effect on the current account balance.

The final long-run effect of an increase in the U.S. money supply in a floating exchange rate system is a depreciation of the U.S. dollar and no change in real GNP. Along the way, GNP temporarily rises and unemployment falls below the natural rate. However, this spurs an increase in the price level, which reduces GNP to its full employment level and raises unemployment back to its natural rate. U.S. inflation occurs in the transition while the price level is increasing.

KEY TAKEAWAY

- The final long-run effect of an increase in the money supply in a floating exchange rate system is a depreciation of the currency and no change in real GNP. In the transition process, there is an inflationary effect.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP in the long run if the nominal money supply increases in the AA-DD model with floating exchange rates.
 - b. Of *increase, decrease, or stay the same*, this is the effect on the domestic currency value in the long run if the nominal money supply increases in the AA-DD model with floating exchange rates.
 - c. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP in the long run if the nominal money supply decreases in the AA-DD model with floating exchange rates.
 - d. Of *increase, decrease, or stay the same*, this is the effect on the domestic currency value in the long run if the nominal money supply decreases in the AA-DD model with floating exchange rates.
2. Repeat the analysis in the text for contractionary monetary policy. Explain each of the four adjustment steps and depict them on an AA-DD diagram.

21.5 Foreign Exchange Interventions with Floating Exchange Rates

LEARNING OBJECTIVES

1. Learn how a country's central bank can intervene to affect the value of the country's currency in a floating exchange rate system.
2. Learn the mechanism and purpose of a central bank sterilized intervention in a Forex market.

In a pure floating exchange rate system, the exchange rate is determined as the rate that equalizes private market demand for a currency with private market supply. The central bank has no necessary role to play in the determination of a pure floating exchange rate. Nonetheless, sometimes central banks desire or are pressured by external groups to take actions (i.e., intervene) to either raise or lower the exchange rate in a floating exchange system. When central banks do intervene on a semiregular basis, the system is sometimes referred to as a "dirty float." There are several reasons such interventions occur.

The first reason central banks intervene is to stabilize fluctuations in the exchange rate. International trade and investment decisions are much more difficult to make if the exchange rate value is changing rapidly. Whether a trade deal or international investment is good or bad often depends on the value of the exchange rate that will prevail at some point in the future. (See [Chapter 15 "Foreign Exchange Markets and Rates of Return"](#), [Section 15.3 "Calculating Rate of Returns on International Investments"](#) for a discussion of how future exchange rates affect returns on international investments.) If the exchange rate changes rapidly, up or down, traders and investors will become more uncertain about the profitability of trades and investments and will likely reduce their international activities. As a consequence, international traders and investors tend to prefer more stable exchange rates and will often pressure governments and central banks to intervene in the foreign exchange (Forex) market whenever the exchange rate changes too rapidly.

The second reason central banks intervene is to reverse the growth in the country's trade deficit. Trade deficits (or current account deficits) can rise rapidly if a country's exchange rate appreciates significantly. A higher currency value will make foreign goods and services (G&S) relatively cheaper, stimulating imports, while domestic goods will seem relatively more expensive to foreigners, thus reducing exports. This means a rising currency value can lead to a rising trade

deficit. If that trade deficit is viewed as a problem for the economy, the central bank may be pressured to intervene to reduce the value of the currency in the Forex market and thereby reverse the rising trade deficit.

There are two methods central banks can use to affect the exchange rate. The indirect method is to change the domestic money supply. The direct method is to intervene directly in the foreign exchange market by buying or selling currency.

Indirect Forex Intervention

The central bank can use an indirect method to raise or lower the exchange rate through domestic money supply changes. As was shown in [Chapter 21 "Policy Effects with Floating Exchange Rates"](#), [Section 21.2 "Monetary Policy with Floating Exchange Rates"](#), increases in the domestic U.S. money supply will cause an increase in $E_{\$/£}$, or a dollar depreciation. Similarly, a decrease in the money supply will cause a dollar appreciation.

Despite relatively quick adjustments in assets markets, this type of intervention must traverse from open market operations to changes in domestic money supply, domestic interest rates, and exchange rates due to new rates of returns. Thus this method may take several weeks or more for the effect on exchange rates to be realized.

A second problem with this method is that to affect the exchange rate the central bank must change the domestic interest rate. Most of the time, central banks use interest rates to maintain stability in domestic markets. If the domestic economy is growing rapidly and inflation is beginning to rise, the central bank may lower the money supply to raise interest rates and help slow down the economy. If the economy is growing too slowly, the central bank may raise the money supply to lower interest rates and help spur domestic expansion. Thus to change the exchange rate using the indirect method, the central bank may need to change interest rates away from what it views as appropriate for domestic concerns at the moment. (Below we'll discuss the method central banks use to avoid this dilemma.)

Direct Forex Intervention

The most obvious and direct way for central banks to intervene and affect the exchange rate is to enter the private Forex market directly by buying or selling domestic currency. There are two possible transactions.

First, the central bank can sell domestic currency (let's use dollars) in exchange for a foreign currency (say, pounds). This transaction will raise the supply of dollars on the Forex (also raising the demand for pounds), causing a reduction in the value of the dollar and thus a dollar depreciation. Of course, when the dollar depreciates in value, the pound appreciates in value with respect to the dollar. Since the central bank is the ultimate source of all dollars (it can effectively print an unlimited amount), it can flood the Forex market with as many dollars as it desires. Thus the central bank's power to reduce the dollar value by direct intervention in the Forex is virtually unlimited.

If instead, the central bank wishes to raise the value of the dollar, it will have to reverse the transaction described above. Instead of selling dollars, it will need to buy dollars in exchange for pounds. The increased demand for dollars on the Forex by the central bank will raise the value of the dollar, thus causing a dollar appreciation. At the same time, the increased supply of pounds on the Forex explains why the pound will depreciate with respect to the dollar.

The ability of a central bank to raise the value of its currency through direct Forex interventions is limited, however. In order for the U.S. Federal Reserve Bank (or the Fed) to buy dollars in exchange for pounds, it must have a stockpile of pound currency (or other pound assets) available to exchange. Such holdings of foreign assets by a central bank are called **foreign exchange reserves**⁸. Foreign exchange reserves are typically accumulated over time and held in case an intervention is desired. In the end, the degree to which the Fed can raise the dollar value with respect to the pound through direct Forex intervention will depend on the size of its pound denominated foreign exchange reserves.

Indirect Effect of Direct Forex Intervention

There is a secondary indirect effect that occurs when a central bank intervenes in the Forex market. Suppose the Fed sells dollars in exchange for pounds in the private Forex. This transaction involves a purchase of foreign assets (pounds) in exchange for U.S. currency. Since the Fed is the ultimate source of dollar currency, these dollars used in the transaction will enter into circulation in the economy in precisely the same way as new dollars enter when the Fed buys a Treasury bill on the open market. The only difference is that with an open market operation, the Fed purchases a domestic asset, while in the Forex intervention it buys a foreign asset. But both are assets all the same and both are paid for with newly created money. Thus when the Fed buys pounds and sells dollars on the Forex, there will be an increase in the U.S. money supply.

8. Holdings of foreign assets by a country's central bank. Usually held in the form of foreign government Treasury bonds.

The higher U.S. money supply will lower U.S. interest rates, reduce the rate of return on U.S. assets as viewed by international investors, and result in a depreciation of the dollar. The direction of this indirect effect is the same as the direct effect.

In contrast, if the Fed were to buy dollars and sell pounds on the Forex, there will be a decrease in the U.S. money supply. The lower U.S. money supply will raise U.S. interest rates, increase the rate of return on U.S. assets as viewed by international investors, and result in an appreciation of the dollar.

The only difference between the direct and indirect effects is the timing and sustainability. The direct effect will occur immediately with central bank intervention since the Fed will be affecting today's supply of dollars or pounds on the Forex. The indirect effect, working through money supply and interest rates, may take several days or weeks. The sustainability of the direct versus indirect effects is discussed next when we introduce the idea of a sterilized Forex intervention.

Sterilized Forex Interventions

There are many times in which a central bank either wants or is pressured to affect the exchange rate value by intervening directly in the foreign exchange market. However, as shown above, direct Forex interventions will change the domestic money supply. A change in the money supply will affect the average interest rate in the short run and the price level, and hence the inflation rate, in the long run. Because central banks are generally entrusted to maintain domestic price stability or to assist in maintaining appropriate interest rates, a low unemployment rate, and GDP growth, Forex intervention will often interfere with one or more of their other goals.

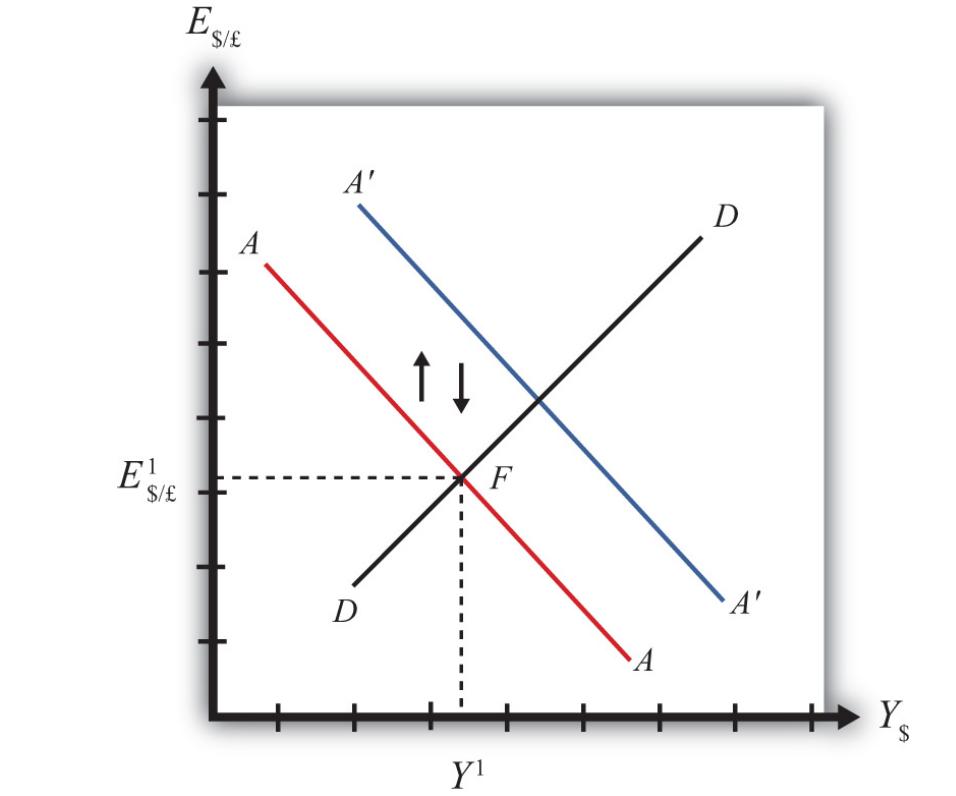
For example, if the central bank believes that current interest rates should be raised slowly during the next several months to slow the growth of the economy and prevent a resurgence of inflation, then a Forex intervention to lower the value of the domestic currency would result in increases in the money supply and a decrease in interest rates, precisely the opposite of what the central bank wants to achieve. Conflicts such as this one are typical and usually result in a central bank choosing to sterilize its Forex interventions.

The intended purpose of a sterilized intervention is to cause a change in the exchange rate while at the same time leaving the money supply and hence interest rates unaffected. As we will see, the intended purpose is unlikely to be realized in practice.

A sterilized foreign exchange intervention occurs when a central bank counters direct intervention in the Forex with a simultaneous offsetting transaction in the domestic bond market. For example, suppose the U.S. Fed decides to intervene to lower the value of the U.S. dollar. This would require the Fed to sell dollars and buy foreign currency on the Forex. Sterilization, in this case, involves a Fed open market operation in which it sells Treasury bonds (T-bonds) at the same time and in the same value as the dollar sale in the Forex market. For example, if the Fed intervenes and sells \$10 million on the Forex, sterilization means it will also sell \$10 million of Treasury bonds on the domestic open market at the same time.

Consider the effects of a sterilized Forex intervention by the U.S. Fed shown in the adjoining AA-DD diagram, [Figure 21.5 "Sterilization in the AA-DD Model"](#). Suppose the economy is initially in equilibrium at point F with GDP (Y^1) and exchange rate (E/£^1$). Now, suppose the Fed intervenes in the Forex by selling dollars and buying British pounds. The direct effect on the exchange rate is not represented in the AA-DD diagram. The only way it can have an effect is through the increase in the money supply, which will shift the AA curve up from AA to A'A'. However, sterilization means the Fed will simultaneously conduct an offsetting open market operation, in this case selling Treasury bonds equal in value to the Forex sales. The sale of T-bonds will lower the U.S. money supply, causing an immediate shift of the AA curve back from A'A' to AA. In fact, because the two actions take place on the same day or within the same week at least, the AA curve does not really shift out at all. Instead, a sterilized Forex intervention maintains the U.S. money supply and thus achieves the Fed's objective of maintaining interest rates.

Figure 21.5 *Sterilization in the AA-DD Model*



However, because there is no shift in the AA or DD curves, the equilibrium in the economy will never move away from point F. This implies that a sterilized Forex intervention not only will not affect GNP, but also will not affect the exchange rate. This suggests the impossibility of the Fed's overall objective to lower the dollar value while maintaining interest rates.

Empirical studies of the effects of sterilized Forex interventions tend to support the results of this simple model. In other words, real-world sterilizations have generally been ineffective in achieving any lasting effect upon a country's currency value.

However, there are several reasons why sterilized interventions may be somewhat effective nonetheless. Temporary effects are certainly possible. If a central bank makes a substantial intervention in the Forex over a short period, this will certainly change the supply or demand of currency and have an immediate effect on the exchange rate on those days.

A more lasting impact can occur if the intervention leads investors to change their expectations about the future. This could happen if investors are not sure whether

the central bank is sterilizing its interventions. Knowing that sterilization is occurring would require a careful observation of several markets unless the Fed announces its policy. However, rather than announcing a sterilized intervention, a central bank that wants to affect expectations should announce the Forex intervention while hiding its offsetting open market operation. In this way, investors may be fooled into thinking that the Forex intervention will lower the future dollar value and thus may adjust their expectations.

If investors are fooled, they will raise $E_{\$/\text{£}}^e$ in anticipation of the future dollar depreciation. The increase in $E_{\$/\text{£}}^e$ will shift the AA curve upward, resulting in an increase in GNP and a depreciation of the dollar. In this way, sterilized interventions may have a more lasting effect on the exchange rate. However, the magnitude of the exchange rate change in this case—if it occurs—will certainly be less than that achieved with a nonsterilized intervention.

KEY TAKEAWAYS

- If the central bank sells domestic currency in exchange for a foreign currency on the Forex, it will cause a direct reduction in the value of the domestic currency, or a currency depreciation.
- If the Fed were to sell dollars on the Forex, there will be an increase in the U.S. money supply that will reduce U.S. interest rates, decrease the rate of return on U.S. assets, and lead to a depreciation of the dollar.
- A sterilized foreign exchange intervention occurs when a central bank counters direct intervention in the Forex with a simultaneous offsetting transaction in the domestic bond market.
- The intended purpose of a sterilized intervention is to cause a change in the exchange rate while at the same time leaving interest rates unaffected.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *buy domestic currency or sell domestic currency* on the foreign exchange market, this is one thing a central bank can do to cause a domestic currency depreciation.
 - b. Of *buy foreign currency or sell foreign currency* on the foreign exchange market, this is one thing a central bank can do to cause a domestic currency appreciation.
 - c. Of *increase, decrease, or keep the same*, this is one thing a central bank can do to the domestic money supply to induce a domestic currency appreciation.
 - d. Of *increase, decrease, or keep the same*, this is one thing a central bank can do to the domestic money supply to induce a domestic currency depreciation.
 - e. The term used to describe a central bank transaction on the domestic bond market intended to offset the central bank’s intervention on the foreign exchange market.
 - f. Of *increase, decrease, or stay the same*, this is the effect on equilibrium GNP in the short run if the central bank sterilizes a sale of foreign reserves on the foreign exchange market in the AA-DD model with floating exchange rates.
 - g. Of *increase, decrease, or stay the same*, this is the effect on the domestic currency value in the short run if the central bank sterilizes a purchase of foreign reserves on the foreign exchange market in the AA-DD model with floating exchange rates.

Chapter 22

Fixed Exchange Rates

Fixed exchange rates around the world were once the only game in town; however, since the collapse of the Bretton Woods system in 1973, floating exchange rates predominate for the world's most-traded currencies. Nonetheless, many countries continue to use some variant of fixed exchange rates even today. This chapter addresses both the historical fixed exchange rate systems like the gold standard as well as the more modern variants like crawling pegs and currency boards.

22.1 Overview of Fixed Exchange Rates

LEARNING OBJECTIVE

1. Preview the discussion about fixed exchange rate systems, their varieties, and their mechanisms.

This chapter begins by defining several types of fixed exchange rate systems, including the gold standard, the reserve currency standard, and the gold exchange standard. The **price-specie flow mechanism**¹ is described for the gold standard. It continues with other modern fixed exchange variations such as fixing a currency to a basket of several other currencies, crawling pegs, fixing within a band or range of exchange rates, currency boards, and finally the most extreme way to fix a currency: adopting another country's currency as your own, as is done with **dollarization**² or euroization.

1. A description about how adjustments to shocks or changes are handled within a pure gold standard system.
2. Currency fixing by adopting the U.S. dollar as one's currency.
3. A currency standard in which all countries fix to one central reserve currency, while the reserve currency is not fixed to anything.
4. The balance on the balance of payments when the central bank sells domestic currency and buys foreign currency.
5. The balance on the balance of payments when the central bank buys domestic currency and sells foreign reserves.
6. A currency standard in which currency is fixed to a weight of gold, and the central bank freely exchanges gold for currency with the public.

The chapter proceeds with the basic mechanics of a **reserve currency standard**³ in which a country fixes its currency to another's. In general, a country's central bank must intervene in the foreign exchange (Forex) markets, buying foreign currency whenever there is excess supply (resulting in a **balance of payments surplus**⁴) and selling foreign currency whenever there is excess demand (resulting in a **balance of payments deficit**⁵). These actions will achieve the fixed exchange rate version of the interest parity condition in which interest rates are equalized across countries. However, to make central bank actions possible, a country will need to hold a stock of foreign exchange reserves. If a country's central bank does not intervene in the Forex in a fixed exchange system, black markets are shown to be a likely consequence.

Results

- **Gold standard**⁶ rules: (1) fix currency to a weight of gold; (2) central bank freely exchanges gold for currency with public.
- Adjustment under a gold standard involves the flow of gold between countries, resulting in equalization of prices satisfying purchasing power parity (PPP) and/or equalization of rates of return on assets satisfying interest rate parity (IRP) at the current fixed exchange rate.
- Reserve currency rules: (1) fix currency to another currency, known as the reserve currency; (2) central bank must hold a stock of foreign exchange reserves to facilitate Forex interventions.

- Gold-exchange standard rules: (1) reserve country fixes its currency to a weight of gold, (2) all other countries fix their currencies to the reserve, (3) reserve central bank freely exchanges gold for currency with other central banks, (4) nonreserve countries hold a stock of the reserve currency to facilitate intervention in the Forex.
- The post–World War II fixed exchange rate system, known as the Bretton Woods system, was a gold exchange standard.
- Some countries fix their currencies to a weighted average of several other currencies, called a “basket of currencies.”
- Some countries implement a crawling peg in which the fixed exchange rate is adjusted regularly.
- Some countries set a central exchange rate and allow free floating within a predefined range or band.
- Some countries implement currency boards to legally mandate Forex interventions.
- Some countries simply adopt another country’s currency, as with dollarization, or choose a brand-new currency, as with the euro.
- The interest rate parity condition becomes the equalization of interest rates between two countries in a fixed exchange rate system.
- A balance of payments surplus (deficit) arises when the central bank buys (sells) foreign reserves on the Forex in exchange for its own currency.
- A black market in currency trade arises when there is unsatisfied excess demand or supply of foreign currency in exchange for domestic currency on the Forex.

KEY TAKEAWAY

- See the main results previewed above.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term for the currency standard that fixes its circulating currency to a quantity of gold.
 - b. The term for the currency standard in which a reserve currency is fixed to a quantity of gold while all other currencies are fixed to the reserve currency.
 - c. The currency standard used during the post–World War II Bretton Woods era.
 - d. The term describing the deficits and surpluses run by a country to maintain a fixed exchange rate.
 - e. The term used to describe a decision by another country to adopt the U.S. dollar as its currency.
 - f. The nonintervention in the Forex market by a country’s central bank is likely to lead to the development of these kinds of market activities.

22.2 Fixed Exchange Rate Systems

LEARNING OBJECTIVES

1. Recognize the varieties of ways that exchange rates can be fixed to a particular value.
2. Understand the basic operation and the adjustment mechanism of a gold standard.

There are two basic systems that can be used to determine the exchange rate between one country's currency and another's: a floating exchange rate system and a fixed exchange rate system.

Under a floating exchange rate system, the value of a country's currency is determined by the supply and demand for that currency in exchange for another in a private market operated by major international banks.

In contrast, in a fixed exchange rate system, a country's government announces (or decrees) what its currency will be worth in terms of *something else* and also sets up the *rules of exchange*. The “*something else*” to which a currency value is set and the “*rules of exchange*” determine the type of fixed exchange rate system, of which there are many. For example, if the government sets its currency value in terms of a fixed weight of gold, then we have a gold standard. If the currency value is set to a fixed amount of another country's currency, then it is a reserve currency standard.

As we review several ways in which a fixed exchange rate system can work, we will highlight some of the advantages and disadvantages of the system. In anticipation, it is worth noting that one key advantage of fixed exchange rates is the intention to eliminate exchange rate risk, which can greatly enhance international trade and investment. A second key advantage is the discipline a fixed exchange rate system imposes on a country's monetary authority, with the intention of inducing a much lower inflation rate.

The Gold Standard

Most people are aware that at one time the world operated under something called a gold standard. Some people today, reflecting back on the periods of rapid growth and prosperity that occurred when the world was on a gold standard, have

suggested that the world abandon its current mixture of fixed and floating exchange rate systems and return to this system. (For a discussion of some pros and cons see [Alan Greenspan's remarks](#) on this from the early 1980s. See Alan Greenspan's remarks in "Can the US Return to a Gold Standard," *Wall Street Journal*, September 1, 1981; reprinted online at [http://www.gold-eagle.com/greenspan011098.html\[0\]](http://www.gold-eagle.com/greenspan011098.html[0]). See [Murray Rothbard's article](#) for an argument in favor of a return to the gold standard. See Murray Rothbard, "The Case for a Genuine Gold Dollar," in *The Gold Standard: An Austrian Perspective* (Lexington, MA: D. C. Heath, 1985), 1–17; also available online at <http://www.mises.org/rothbard/genuine.asp>.) Whether or not countries seriously consider this in the future, it is instructive to understand the workings of a gold standard, especially since, historically, it is the first major international system of fixed exchange rates.

Most of the world maintained a pure gold standard during the late 1800s and early 1900s, with a major interruption during World War I. Before the enactment of a gold standard, countries were generally using a [Bimetallic standard](#)⁷ consisting of both gold and silver. See Angela Radish, "Bimetallism," Economic History Online at <http://www.eh.net/encyclopedia/?article=redish.bimetallism>. The earliest establishment of a gold standard was in Great Britain in 1821, followed by Australia in 1852 and Canada in 1853. The United States established its gold standard system with the Coinage Act of 1873, sometimes known as "[The Crime of '73](#)." For more info see Wikipedia, "Coinage Act of 1873," http://en.wikipedia.org/wiki/Coinage_Act_of_1873. The gold standard was abandoned in the early days of the Great Depression. Britain dropped the standard in 1931, the United States in 1933.

The rules of a gold standard are quite simple. First, a country's government declares that its issued currency (it may be coin or paper currency) will exchange for a weight in gold. For example, in the United States during the late 1800s and early 1900s, the government set the dollar exchange rate to gold at the rate \$20.67 per troy ounce. During the same period, Great Britain set its currency at the rate £4.24 per troy ounce. Second, in a pure gold standard, a country's government declares that it will freely exchange currency for actual gold at the designated exchange rate. This "rule of exchange" means that anyone can go to the central bank with coin or currency and walk out with pure gold. Conversely, one could also walk in with pure gold and walk out with the equivalent in coin or currency.

7. A currency standard using both gold and silver.

8. Gold held in storage by a central bank, usually to make possible the exchange of currency for gold under a gold standard.

Because the government bank must always be prepared to give out gold in exchange for coin and currency on demand, it must maintain a storehouse of gold. That store of gold is referred to as "**gold reserves**". That is, the central bank maintains a reserve of gold so that it can always fulfill its promise of exchange. As discussed in [Chapter 22 "Fixed Exchange Rates"](#), [Section 22.4 "Central Bank Intervention with Fixed Exchange Rates"](#), a well-functioning system will require that the central bank always have an adequate amount of reserves.

The two simple rules, when maintained, guarantee that the exchange rate between dollars and pounds remains constant. Here's why.

First, the dollar/pound exchange rate is defined as the ratio of the two-currency-gold exchange rates. Thus

$$E_{\$/\text{£}} = \frac{20.67 \text{ \$/oz}}{4.24 \text{ £/oz}} = 4.875 \frac{\$}{\text{oz}} \frac{\text{oz}}{\text{£}} = 4.875 \frac{\$}{\text{£}}.$$

Next, suppose an individual wants to exchange \$4.875 for one pound. Following the exchange rules, this person can enter the central bank in the United States and exchange dollars for gold to get

$$\frac{\$4.875}{20.67 \text{ \$/oz}} = 0.23585 \text{ oz of gold.}$$

This person can then take the gold into the central bank in the United Kingdom, and assuming no costs of transportation, can exchange the gold into pounds as follows:

$$0.23585 \text{ oz} \times 4.24 \frac{\text{£}}{\text{oz}} = \text{£}1.00.$$

Hence, the \$4.875 converts to precisely £1 and this will remain the fixed exchange rate between the two currencies, as long as the simple exchange rules are followed. If many countries define the value of their own currency in terms of a weight of gold and agree to exchange gold freely at that rate with all who desire to exchange, then all these countries will have fixed currency exchange rates with respect to each other.

Price-Specie Flow Mechanism

The price-specie flow mechanism is a description about how adjustments to shocks or changes are handled within a pure gold standard system. Although there is some disagreement whether the gold standard functioned as described by this mechanism, the mechanism does fix the basic principles of how a gold standard is supposed to work.

Consider the United States and United Kingdom operating under a pure gold standard. Suppose there is a gold discovery in the United States. This will represent a shock to the system. Under a gold standard, a gold discovery is like digging up

money, which is precisely what inspired so many people to *rush* to California after 1848 to strike it rich.

Once the gold is unearthed, the prospectors bring it into town and proceed to the national bank where it can be exchanged for coin and currency at the prevailing dollar/gold exchange rate. The new currency in circulation represents an increase in the domestic money supply.

Indeed, it is this very transaction that explains the origins of the gold and silver standards in the first place. The original purpose of banks was to store individuals' precious metal wealth and to provide exchangeable notes that were backed by the gold holdings in the vault. Thus rather than carrying around heavy gold, one could carry lightweight paper money. Before national or central banks were founded, individual commercial banks issued their own currencies, which circulated together with many other bank currencies. However, it was also common for governments to issue coins that were made directly from gold or silver.

Now, once the money supply increases following the gold discovery, it can have two effects: operating through the goods market and financial market. The price-specie flow mechanism describes the adjustment through goods markets.

First, let's assume that the money increase occurs in an economy that is not growing—that is, with a fixed level of GDP. Also assume that both purchasing power parity (PPP) and interest rate parity (IRP) holds. PPP implies an equalization of the cost of a market basket of goods between the United States and the United Kingdom at the current fixed exchange rate. IRP implies an equalization of the rates of return on comparable assets in the two countries.

As discussed in [Chapter 18 "Interest Rate Determination"](#), [Section 18.14 "Money Supply and Long-Run Prices"](#), when the U.S. money supply increases, and when there is no subsequent increase in output, the prices of goods and services will begin to rise. This inflationary effect occurs because more money is chasing (i.e., demanding) the same amount of goods and services. As the price level rises in an economy open to international trade, domestic goods become more expensive relative to foreign goods. This will induce domestic residents to increase demand for foreign goods; hence import demand will rise. Foreign consumers will also find domestic goods more expensive, so export supply will fall. The result is a demand for a current account deficit. To make these transactions possible in a gold standard, currency exchange will take place as follows.

U.S. residents wishing to buy cheaper British goods will first exchange dollars for gold at the U.S. central bank. Then they will ship that gold to the United Kingdom to exchange for the pounds that can be used to buy UK goods. As gold moves from the United States to the United Kingdom, the money supply in the United States falls while the money supply in the United Kingdom rises. Less money in the United States will eventually reduce prices, while more money in the United Kingdom will raise prices. This means that the prices of goods will move together until purchasing power parity holds again. Once PPP holds, there is no further incentive for money to move between countries. There will continue to be demand for UK goods by U.S. residents, but this will balance with the United Kingdom demands for similarly priced U.S. goods. Hence, the trade balance reverts to zero.

The adjustment process in the financial market under a gold standard will work through changes in interest rates. When the U.S. money supply rises after the gold discovery, average interest rates will begin to fall. Lower U.S. interest rates will make British assets temporarily more attractive, and U.S. investors will seek to move investments to the United Kingdom. The adjustment under a gold standard is the same as with goods. Investors trade dollars for gold in the United States and move that gold to the United Kingdom where it is exchanged for pounds and used to purchase UK assets. Thus the U.S. money supply will begin to fall, causing an increase in U.S. interest rates, while the UK money supply rises, leading to a decrease in UK interest rates. The interest rates will move together until interest rate parity again holds.

In summary, *adjustment under a gold standard involves the flow of gold between countries, resulting in equalization of prices satisfying purchasing power parity (PPP) and/or equalization of rates of return on assets satisfying interest rate parity (IRP) at the current fixed exchange rate.* The only requirement for the government to maintain this type of fixed exchange rate system is to maintain the fixed price of its currency in terms of gold and to freely and readily exchange currency for gold on demand.

Reserve Currency Standard

In a reserve currency system, another country's currency takes the role that gold played in a gold standard. In other words a country fixes its own currency value to a unit of another country's currency. For example, suppose Britain decided to fix its currency to the dollar at the exchange rate $E\$/\text{£} = 1.50$. To maintain this fixed exchange rate, the Bank of England would stand ready to exchange pounds for dollars (or dollars for pounds) on demand at the specified exchange rate. To accomplish this, the Bank of England would need to hold dollars *on reserve* in case there was ever any excess demand for dollars in exchange for pounds on the Forex. In the gold standard, the central bank held gold to exchange for its own currency;

with a reserve currency standard, it must hold a stock of the reserve currency. Always, the reserve currency is the currency to which the country fixes.

A reserve currency standard is the typical method for fixing a currency today. Most countries that fix its exchange rate will fix to a currency that either is prominently used in international transactions or is the currency of a major trading partner. Thus many countries fixing their exchange rate today fix to the U.S. dollar because it is the most widely traded currency internationally. Alternatively, fourteen African countries that were former French colonies had established the CFA franc zone and fixed the CFA franc (current currency used by these African countries) to the French franc. Since 1999, the CFA franc has been fixed to the euro. Namibia, Lesotho, and Swaziland are all a part of the common monetary area (CMA) and fix their currency to the South African rand.

Gold Exchange Standard

A **gold exchange standard**⁹ is a mixed system consisting of a cross between a reserve currency standard and a gold standard. In general, it includes the following two rules:

1. A reserve currency is chosen. All nonreserve countries agree to fix their exchange rates to the reserve at some announced rate. To maintain the fixity, these nonreserve countries will hold a stockpile of reserve currency assets.
2. The reserve currency country agrees to fix its currency value to a weight in gold. Finally, the reserve country agrees to exchange gold for its own currency with other central banks within the system on demand.

One key difference in this system from a gold standard is that the reserve country does not agree to exchange gold for currency with the general public, only with other central banks.

The system works exactly like a reserve currency system from the perspective of the nonreserve countries. However, if over time the nonreserve countries accumulate the reserve currency, they can demand exchange for gold from the reserve country central bank. In this case, gold reserves will flow away from the reserve currency country.

9. When all countries fix to one central reserve currency, while the reserve currency is fixed to gold.

The fixed exchange rate system set up after World War II was a gold exchange standard, as was the system that prevailed between 1920 and the early 1930s. The

post–World War II system was agreed to by the allied countries at a conference in Bretton Woods, New Hampshire, in the United States in June 1944. As a result, the exchange rate system after the war also became known as the **Bretton Woods system**¹⁰.

Also proposed at Bretton Woods was the establishment of an international institution to help regulate the fixed exchange rate system. This institution was the **International Monetary Fund (IMF)**¹¹. The IMF's main mission was to help maintain the stability of the Bretton Woods fixed exchange rate system.

Other Fixed Exchange Rate Variations

Basket of Currencies

Countries that have several important trading partners, or who fear that one currency may be too volatile over an extended period, have chosen to fix their currency to a basket of several other currencies. This means fixing to a weighted average of several currencies. This method is best understood by considering the creation of a composite currency. Consider the following hypothetical example: a new unit of money consisting of 1 euro, 100 Japanese yen, and one U.S. dollar. Call this new unit a Eur-yen-dol. A country could now fix its currency to one Eur-yen-dol. The country would then need to maintain reserves in one or more of the three currencies to satisfy excess demand or supply of its currency on the Forex.

A better example of a composite currency is found in the SDR. SDR stands for **special drawing rights**¹². It is a composite currency created by the International Monetary Fund (IMF). One SDR now consists of a fixed quantity of U.S. dollars, euros, Japanese yen, and British pounds. For more info on the SDR see the [IMF factsheet](#). International Monetary Fund, About the IMF, Factsheet, “Special Drawing Rights (SDRs),” [http://www.imf.org/external/np/exr/facts/sdr.htm\[0\]](http://www.imf.org/external/np/exr/facts/sdr.htm[0]). Now Saudi Arabia officially fixes its currency to the SDR. Botswana fixes to a basket consisting of the SDR and the South African rand.

10. The fixed exchange rate system (using a gold exchange standard) set up after World War II and lasting until 1973.
11. The international organization created after World War II to oversee the Bretton Woods system of fixed exchange rates.
12. A composite currency created by the International Monetary Fund (IMF), used only for transactions between central banks.

Crawling Pegs

A crawling peg refers to a system in which a country fixes its exchange rate but also changes the fixed rate at periodic or regular intervals. Central bank interventions in the Forex may occur to maintain the temporary fixed rate. However, central banks can avoid interventions and save reserves by adjusting the fixed rate instead. Since crawling pegs are adjusted gradually, they can help eliminate some exchange rate volatility without fully constraining the central bank with a fixed rate. In 2010 Bolivia, China, Ethiopia, and Nicaragua were among several countries maintaining a crawling peg.

Pegged within a Band

In this system, a country specifies a central exchange rate together with a percentage allowable deviation, expressed as plus or minus some percentage. For example, Denmark, an EU member country, does not yet use the euro but participates in the Exchange Rate Mechanism (ERM2). Under this system, Denmark sets its central exchange rate to 7.46038 krona per euro and allows fluctuations of the exchange rate within a 2.25 percent band. This means the krona can fluctuate from a low of 7.63 kr/€ to a high of 7.29 kr/€. (Recall that the krona is at a high with the smaller exchange rate value since the kr/euro rate represents the euro value.) If the market determined floating exchange rate rises above or falls below the bands, the Danish central bank must intervene in the Forex. Otherwise, the exchange rate is allowed to fluctuate freely.

As of 2010, Slovenia, Syria, and Tonga were fixing their currencies within a band.

Currency Boards

A currency board is a legislated method to provide greater assurances that an exchange rate fixed to a reserve currency will indeed remain fixed. In this system, the government requires that domestic currency is always exchangeable for the specific reserve at the fixed exchange rate. The central bank authorities are stripped of all discretion in the Forex interventions in this system. As a result, they must maintain sufficient foreign reserves to keep the system intact.

In 2010 Bulgaria, Hong Kong, Estonia, and Lithuania were among the countries using a currency board arrangement. Argentina used a currency board system from 1991 until 2002. The currency board was very effective in reducing inflation in Argentina during the 1990s. However, the collapse of the exchange rate system and the economy in 2002 demonstrated that currency boards are not a panacea.

Dollarization/Euroization

The most extreme and convincing method for a country to fix its exchange rate is to give up one's national currency and adopt the currency of another country. In creating the euro-zone among twelve of the European Union (EU) countries, these European nations have given up their own national currencies and have adopted the currency issued by the European Central Bank. This is a case of euroization. Since all twelve countries now share the euro as a common currency, their exchange rates are effectively fixed to each other at a 1:1 ratio. As other countries in the EU join the common currency, they too will be forever fixing their exchange rate to the euro. (Note, however, that although all countries that use the euro are

fixed to each other, the euro itself floats with respect to external currencies such as the U.S. dollar.)

Other examples of adopting another currency as one's own are the countries of Panama, Ecuador, and El Salvador. These countries have all chosen to adopt the U.S. dollar as their national currency of circulation. Thus they have chosen the most extreme method of assuring a fixed exchange rate. These are examples of dollarization.

KEY TAKEAWAYS

- In a gold standard, a country's government declares that its issued currency will exchange for a weight in gold and that it will freely exchange currency for actual gold at the designated exchange rate.
- Adjustment under a gold standard involves the flow of gold between countries, resulting in equalization of prices satisfying purchasing power parity (PPP) and/or equalization of rates of return on assets satisfying interest rate parity (IRP) at the current fixed exchange rate.
- In a reserve currency system, a country fixes its own currency value to a unit of another country's currency. The other country is called the reserve currency country.
- A gold exchange standard is a mixed system consisting of a cross between a reserve currency standard and a gold standard. First, a reserve currency is chosen. Second, the reserve currency country agrees to fix its currency value to a weight in gold. Finally, the reserve country agrees to exchange gold for its own currency with other central banks within the system on demand.
- The post-World War II Bretton Woods system was a gold exchange currency standard.
- Other fixed exchange rate choices include fixing to a market basket, fixing in a nonrigid way by implementing a crawling peg or an exchange rate band, implementing a currency board, or adopting another country's currency.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe the adjustment mechanism within a gold standard.
 - b. The term given to the currency standard using both gold and silver.
 - c. The term given to the currency standard in which all countries fix to one central currency, while the central currency is not fixed to anything.
 - d. The name of the international organization created after World War II to oversee the fixed exchange rate system.
 - e. In the late nineteenth century, the U.S. dollar was fixed to gold at this exchange rate.
 - f. In the late nineteenth century, the British pound was fixed to gold at this exchange rate.
 - g. In the late nineteenth century, one U.S. dollar was worth approximately this many shillings (note: a shilling is one-tenth of a pound).
 - h. Of *gold inflow* or *gold outflow*, this is likely to occur for a country whose interest rates rise under a gold standard with free capital mobility.
 - i. The term used to describe a currency system in which a country fixes its exchange rate but also changes the fixed rate at periodic or regular intervals.
 - j. As of 2004, Estonia and Hong Kong implemented this type of currency system.
2. Use the IMF’s “De Facto Classification of Exchange Rate Regimes and Monetary Policy Frameworks” at <http://www.imf.org/external/np/mfd/er/2008/eng/0408.htm> to answer the following questions:
 - a. What are four countries that maintained currency board arrangements?
 - b. What are four countries that maintained a conventional fixed peg?

- c. What are four countries that maintained a crawling peg?
- d. What are four countries whose currencies were independently floating?

22.3 Interest Rate Parity with Fixed Exchange Rates

LEARNING OBJECTIVE

1. Learn how the interest rate parity condition changes in a system of credible fixed exchange rates.

One of the main differences between a fixed exchange rate system and a floating system is that under fixed exchange rates the central bank will have to “do something” periodically. In contrast, in a floating system, the central bank can just sit back and watch since it has no responsibility for the value of the exchange rate. In a pure float, the exchange rate is determined entirely by private transactions.

However, in a fixed exchange rate system, the central bank will need to intervene in the foreign exchange market, perhaps daily, if it wishes to maintain the credibility of the exchange rate.

We'll use the AA-DD model to explain why. Although the AA-DD model was created under the assumption of a floating exchange rate, we can reinterpret the model in light of a fixed exchange rate assumption. This means we must look closely at the interest rate parity condition, which represents the equilibrium condition in the foreign exchange market.

Recall that the AA-DD model assumes the exchange rate is determined as a result of investor incentives to maximize their rate of return on investments. The model ignores the potential effect of importers and exporters on the exchange rate value. That is, the model does not presume that purchasing power parity holds. As such, the model describes a world economy that is very open to international capital flows and international borrowing and lending. This is a reasonable representation of the world in the early twenty-first century, but would not be the best characterization of the world in the mid-1900s when capital restrictions were more common. Nonetheless, the requisite behavior of central banks under fixed exchange rates would not differ substantially under either assumption.

When investors seek the greatest rate of return on their investments internationally, we saw that the exchange rate will adjust until interest rate parity holds. Consider interest rate parity (IRP) for a particular investment comparison

between the United States and the United Kingdom. IRP means that $RoR\$ = RoRE$. We can write this equality out in its complete form to get

$$i\$ = i£ + (1 + i£) \frac{E_{\$/£}^e - E_{\$/£}}{E_{\$/£}},$$

where the left-hand side is the U.S. interest rate and the right side is the more complicated rate of return formula for a UK deposit with interest rate $i£$. (See [Chapter 15 "Foreign Exchange Markets and Rates of Return"](#) and [Chapter 16 "Interest Rate Parity"](#) for the derivation of the interest rate parity condition.) The last term on the right represents the expected appreciation (if positive) or depreciation (if negative) of the pound value with respect to the U.S. dollar.

In a floating exchange rate system, the value of this term is based on investor expectations about the future exchange rate as embodied in the term $E_{\$/£}^e$, which determines the degree to which investors believe the exchange rate will change over their investment period.

If these same investors were operating in a fixed exchange rate system, however, and if they believed the fixed exchange rate would indeed remain fixed, then the investors' expected exchange rate should be set equal to the current fixed spot exchange rate. In other words, under credible fixed exchange rates, $E_{\$/£}^e = E_{\$/£}$. Investors should not expect the exchange rate to change from its current fixed value. (We will consider a case in which the investors' expected exchange rate does not equal the fixed spot rate in [Chapter 23 "Policy Effects with Fixed Exchange Rates"](#), [Section 23.6 "Currency Crises and Capital Flight"](#).)

With $E_{\$/£}^e = E_{\$/£}$, the right side of the above expression becomes zero, and the interest rate parity condition under fixed exchange rates becomes

$$i\$ = i£.$$

Thus for interest rate parity to hold in a fixed exchange rate system, the interest rates between two countries must be equal.

Indeed, the reason this condition in a floating system is called “interest rate parity” rather than “rate of return parity” is because of our history with fixed exchange rates. Before 1973, most of the world had maintained fixed exchange rates for most of the time. We can see now that under fixed exchange rates, rates of return in each country are simply the interest rates on individual deposits. In other words, in a

fixed system, which is what most countries had through much of their histories, interest rate parity means the equality of interest rates. When the fixed exchange rate system collapsed, economists and others continued to use the now-outdated terminology: interest rate parity. Inertia in language usage is why the traditional term continues to be applied (somewhat inappropriately) even today.

KEY TAKEAWAY

- For interest rate parity to hold in a fixed exchange rate system, the interest rates between two countries must be equal.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. These must be equalized between countries for interest rate parity to hold under fixed exchange.
 - b. If the fixed exchange rates are credible, then the expected exchange rate should be equal to this exchange rate.
 - c. Of *intervene* or *do not intervene*, this is what a central bank should do in the Forex market if it intends to maintain credible fixed exchange rates.

22.4 Central Bank Intervention with Fixed Exchange Rates

LEARNING OBJECTIVE

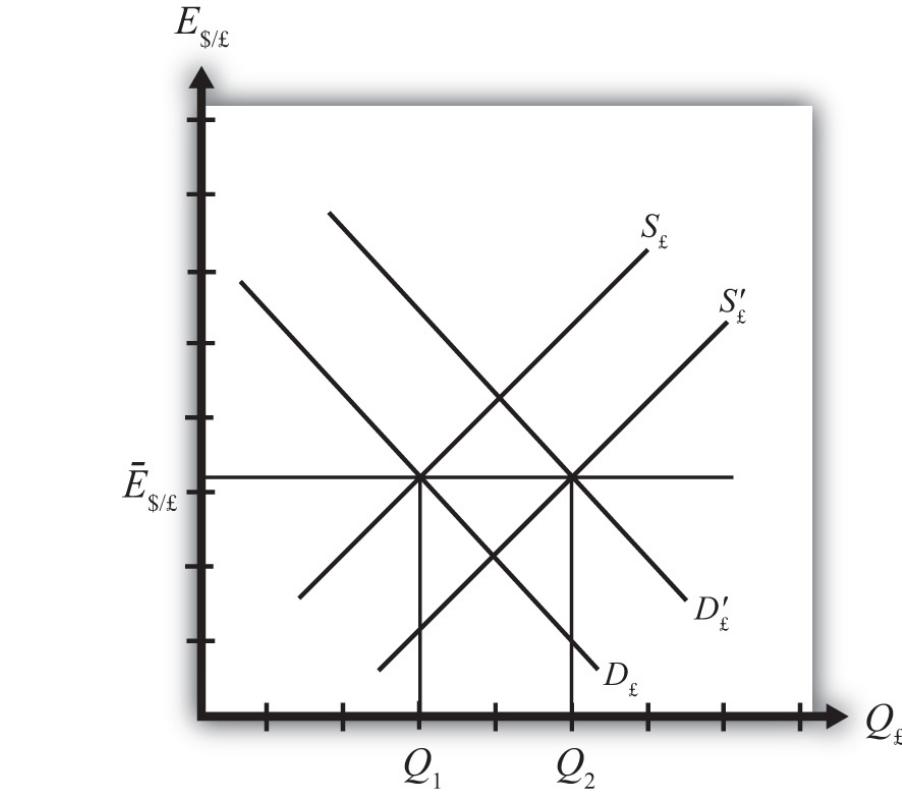
1. Learn what a central bank must do to maintain a credible fixed exchange rate in a reserve currency system.

In a fixed exchange rate system, most of the transactions of one currency for another will take place in the private market among individuals, businesses, and international banks. However, by fixing the exchange rate the government would have declared illegal any transactions that do not occur at the announced rate. However, it is very unlikely that the announced fixed exchange rate will at all times equalize private demand for foreign currency with private supply. In a floating exchange rate system, the exchange rate adjusts to maintain the supply and demand balance. In a fixed exchange rate system, it becomes the responsibility of the central bank to maintain this balance.

The central bank can intervene in the private foreign exchange (Forex) market whenever needed by acting as a buyer and seller of currency of last resort. To see how this works, consider the following example.

Suppose the United States establishes a fixed exchange rate to the British pound at the rate $\bar{E}_{\$/\text{£}}$. In [Figure 22.1 "Central Bank Intervention to Maintain a Fixed Exchange Rate"](#), we depict an initial private market Forex equilibrium in which the supply of pounds (S_E) equals demand (D_E) at the fixed exchange rate ($\bar{E}_{\$/\text{£}}$). But suppose, for some unspecified reason, the demand for pounds on the private Forex rises one day to D'_E .

Figure 22.1 Central Bank Intervention to Maintain a Fixed Exchange Rate

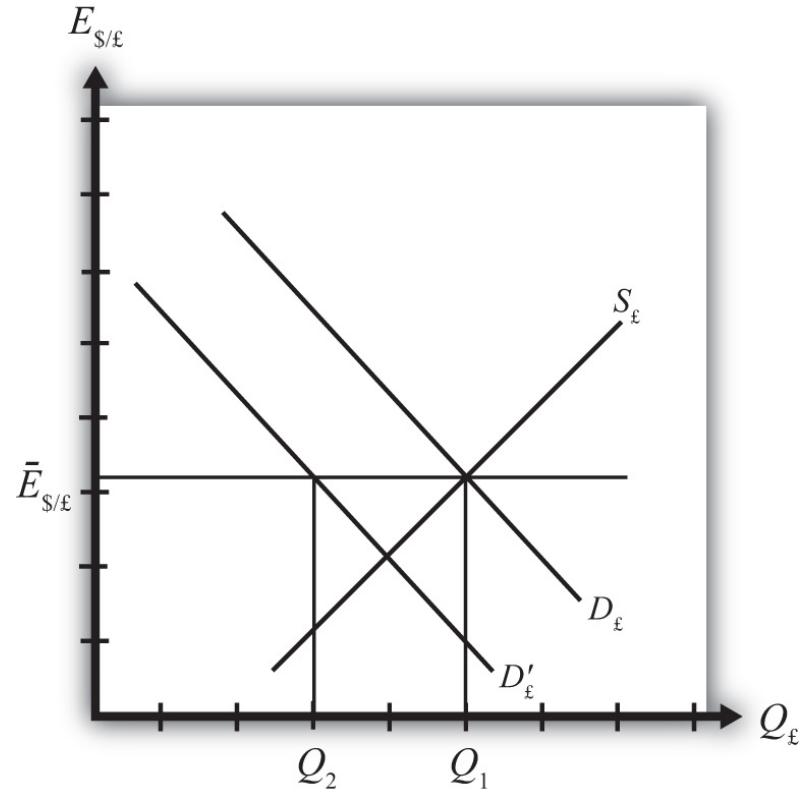


At the fixed exchange rate ($\bar{E}_{\$/\pounds}$), private market demand for pounds is now Q_2 , whereas supply of pounds is Q_1 . This means there is excess demand for pounds in exchange for U.S. dollars on the private Forex.

To maintain a credible fixed exchange rate, the U.S. central bank would immediately satisfy the excess demand by supplying additional pounds to the Forex market. That is, it sells pounds and buys dollars on the private Forex. This would cause a shift of the pound supply curve from S_{\pounds} to S'_{\pounds} . In this way, the equilibrium exchange rate is automatically maintained at the fixed level.

Alternatively, consider [Figure 22.2 "Another Central Bank Intervention to Maintain a Fixed Exchange Rate"](#), in which again the supply of pounds (S_{\pounds}) equals demand (D_{\pounds}) at the fixed exchange rate ($\bar{E}_{\$/\pounds}$). Now suppose, for some unspecified reason, the demand for pounds on the private Forex falls one day to D'_{\pounds} . At the fixed exchange rate ($\bar{E}_{\$/\pounds}$), private market demand for pounds is now Q_2 , whereas supply of pounds is Q_1 . This means there is excess supply of pounds in exchange for U.S. dollars on the private Forex.

Figure 22.2 Another Central Bank Intervention to Maintain a Fixed Exchange Rate



In this case, an excess supply of pounds also means an excess demand for dollars in exchange for pounds. The U.S. central bank can satisfy the extra dollar demand by entering the Forex and selling dollars in exchange for pounds. This means it is supplying more dollars and demanding more pounds. This would cause a shift of the pound demand curve from $D'_{\text{£}}$ back to $D_{\text{£}}$. Since this intervention occurs immediately, the equilibrium exchange rate is automatically and always maintained at the fixed level.

KEY TAKEAWAYS

- If, for example, the United States fixes its currency to the British pound (the reserve), when there is excess demand for pounds in exchange for U.S. dollars on the private Forex, the U.S. central bank would immediately satisfy the excess demand by supplying additional pounds to the Forex market. By doing so, it can maintain a credible fixed exchange rate.
- If, for example, the United States fixes its currency to the British pound (the reserve), when there is excess demand for dollars in exchange for British pounds on the private Forex, the U.S. central bank would immediately satisfy the excess demand by supplying dollars to the Forex market. By doing so, it can maintain a credible fixed exchange rate.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *buy, sell, or do nothing*, this is what a central bank must do with its reserve currency if there is excess demand for its own currency in the private Forex market while maintaining a fixed exchange rate.
 - b. Of *buy, sell, or do nothing*, this is what a central bank must do with its reserve currency if there is excess demand for the reserve currency in the private Forex market while maintaining a fixed exchange rate.
 - c. Of *buy dollars, sell dollars, or do nothing*, this is what China’s central bank must do if there is excess demand for Chinese yuan in the private Forex market if China fixes its currency to the U.S. dollar.
 - d. Of *buy yuan, sell yuan, or do nothing*, this is what China’s central bank must do if there is excess demand for U.S. dollars in the private Forex market if China fixes its currency to the U.S. dollar.

22.5 Balance of Payments Deficits and Surpluses

LEARNING OBJECTIVE

1. Learn the definitions and usage of balance of payments deficits and surpluses in a fixed exchange rate system.

To maintain a fixed exchange rate, the central bank will need to automatically intervene in the private foreign exchange (Forex) by buying or selling domestic currency in exchange for the foreign reserve currency. Clearly, in order for these transactions to be possible, a country's central bank will need a stock of the foreign reserve currency at the time the fixed exchange rate system begins. Subsequently, if excess demand for foreign currency in some periods is balanced with excess supply in other periods, then falling reserves in some periods (when dollars are bought on the Forex) will be offset with rising reserves in other periods (when dollars are sold in the Forex) and a central bank will be able to maintain the fixed exchange rate. Problems arise, though, if a country begins to run out of foreign reserves. But before discussing that situation, we need to explain some terminology.

When the central bank buys domestic currency and sells the foreign reserve currency in the private Forex, the transaction indicates a balance of payments deficit. Alternatively, when the central bank sells domestic currency and buys foreign currency in the Forex, the transaction indicates a balance of payments surplus.

Central bank transactions are recorded in an account titled **official reserve transactions**¹³. It is found in the financial account of the balance of payments. If this account indicates an addition to official reserves over some period, then the country is running a balance of payments surplus. If over some period the official reserve balance is falling, then the country is running a balance of payments deficit. The deficit or surplus terminology arises from the following circumstances.

Suppose a country runs a trade deficit in a fixed exchange rate system. A trade deficit means that demand for imports exceeds foreign demand for our exports. This implies that domestic demand for foreign currency (to buy imports) exceeds foreign demand for domestic currency (to buy our exports). Assuming no additional foreign demands for domestic currency on the financial account (to keep the exchange rate fixed), the central bank would need to intervene by selling foreign currency in exchange for domestic currency. This would lead to a reduction of foreign reserves and hence a balance of payments deficit. In the absence of

13. Account on the balance of payments used to record all central bank transactions.

transactions on the financial account, to have a trade deficit and a fixed exchange rate implies a balance of payments deficit as well.

More generally, a balance of payments deficit (surplus) arises whenever there is excess demand for (supply of) foreign currency on the private Forex at the official fixed exchange rate. To satisfy the excess demand (excess supply), the central bank will automatically intervene on the Forex and sell (buy) foreign reserves. Thus by tracking sales or purchases of foreign reserves in the official reserve account, we can determine if the country has a balance of payments deficit or surplus.

Note that in a floating exchange rate system, a central bank can intervene in the private Forex to push the exchange rate up or down. Thus official reserve transactions can show rising or falling foreign reserves and hence suggest a balance of payments deficit or surplus in a floating exchange system. However, it is not strictly proper to describe a country with floating exchange rates as having a balance of payment deficit or surplus. The reason is that interventions are not *necessary* in a floating exchange rate. In a floating system, an imbalance between supply and demand in the private Forex is relieved by a change in the exchange rate. Thus there need never be an imbalance in the balance of payments in a floating system.

KEY TAKEAWAYS

- When the central bank buys domestic currency and sells the foreign reserve currency in the private Forex, the transaction indicates a balance of payments deficit.
- When the central bank sells domestic currency and buys foreign currency in the Forex, the transaction indicates a balance of payments surplus.
- A balance of payments deficit (surplus) arises whenever there is excess demand for (supply of) foreign currency on the private Forex at the official fixed exchange rate.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The account on the balance of payments (BoP) used to record all central bank transactions.
 - b. The balance on the BoP when the central bank sells foreign reserves.
 - c. Of *BoP deficit*, *BoP surplus*, or *BoP balance*, this is what a central bank will run if there is excess demand for its own currency in the private Forex market while maintaining a fixed exchange rate.
 - d. Of *BoP deficit*, *BoP surplus*, or *BoP balance*, this is what a central bank will run if there is excess demand for the reserve currency in the private Forex market while maintaining a fixed exchange rate.
 - e. Of *BoP deficit*, *BoP surplus*, or *BoP balance*, this is what China’s central bank will run if there is excess demand for Chinese yuan in the private Forex market if China fixes its currency to the U.S. dollar.
 - f. Of *BoP deficit*, *BoP surplus*, or *BoP balance*, this is what China’s central bank will run if there is excess demand for U.S. dollars in the private Forex market if China fixes its currency to the U.S. dollar.

22.6 Black Markets

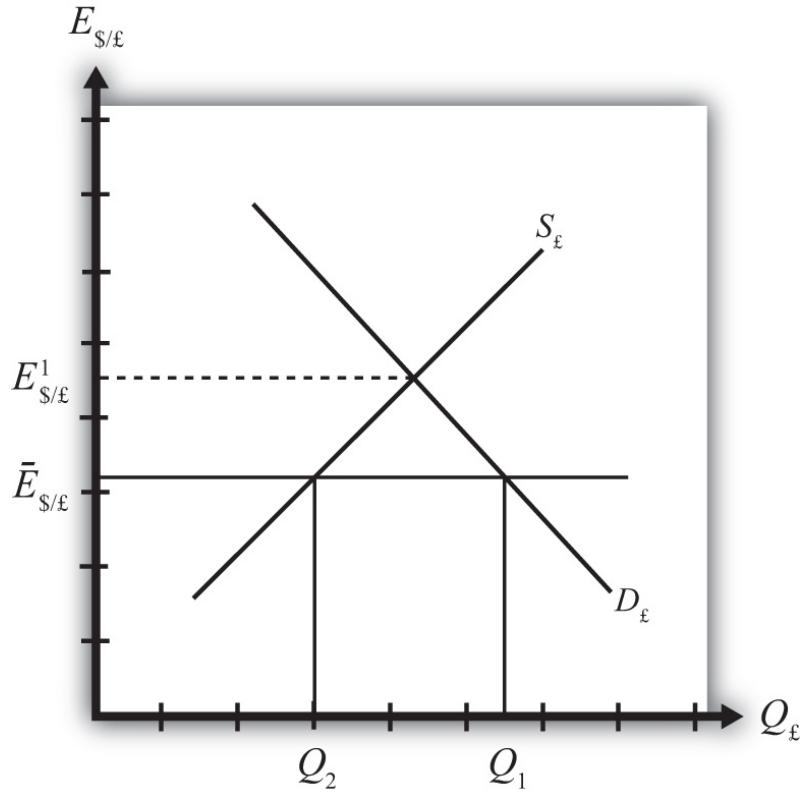
LEARNING OBJECTIVE

1. Learn the five different reasons why trade between countries may occur.

Till now we have said that a central bank must intervene in the foreign exchange (Forex) market whenever there is excess demand or supply of foreign currency. However, we might consider what would happen if the central bank did not intervene. Surely the government could simply mandate that all Forex transactions take place at the official fixed rate and implement severe penalties if anyone is caught trading at a different rate. A black market arises, however, when exchanges for foreign currency take place at an unofficial (or illegal) exchange rate.

Let's consider why a black market may arise. Suppose the United States fixes its exchange rate to the British pound at the rate $\bar{E}_{\$/\text{£}}$. This is indicated in [Figure 22.3 "Conditions for a Black Market"](#) as a horizontal line drawn at $\bar{E}_{\$/\text{£}}$. Suppose further that demand for pounds (Q_1) on the private Forex exceeds supply (Q_2) at the official fixed exchange rate, but the central bank does not intervene to correct the imbalance. In this case, suppliers of pounds will come to the market with Q_2 quantity of pounds, but many people who would like to buy pounds will not find a willing supplier. Those individuals and businesses demanding the excess ($Q_1 - Q_2$) will leave the market empty-handed. Now if this were a one-time occurrence, the unsatisfied demand might be fulfilled in later days when excess supply of pounds comes to the market. However, a more likely scenario is that this unsatisfied demand persists for a long period. With each passing day of unsatisfied demand, total unsatisfied demand grows insidiously.

Figure 22.3 Conditions for a Black Market



Together with the excess demand is a willingness to pay more than the official rate to obtain currency. Since the market equilibrium rate is at $E_{\$/\text{£}}^1$, excess demanders would be willing to pay more dollars to obtain a pound than is required in the official market. The willingness to pay more creates a profit-making possibility. Suppose an individual or business obtains pounds, perhaps by selling goods in Britain and being paid in foreign currency. This person could convert the pounds for dollars at the official rate or, if he or she wants to make more money, could trade the currency “unofficially” at a higher exchange rate. The only problem is finding someone willing to buy the pounds at the unofficial rate. This turns out rarely to be a problem. Wherever black markets develop, unofficial traders find each other on street corners, at hotels, and even within banks.

Thus a central bank doesn't absolutely need to intervene in the Forex market in a fixed exchange rate system. However, if it does not, a black market will very likely arise and the central bank will lose control of the exchange rate. One main purpose of fixed exchange rates, namely the certainty of knowing what the currency will exchange for, is also lost since traders will have to decide whether to trade officially or unofficially. Furthermore, the black market exchange rate typically rises and

falls with changes in supply and demand, thus one is never sure what that rate will be.

In light of the potential for black markets to arise, if a government wishes to maintain a *credible* fixed exchange rate, regular intervention to eliminate excess demand or supply of foreign currency is indeed required.

KEY TAKEAWAYS

- A black market arises when exchanges for foreign currency take place at an unofficial (or illegal) exchange rate.
- If a central bank does not intervene regularly in the Forex market, a black market will very likely arise and the central bank will lose control of the exchange rate.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term used to describe currency transactions that occur at unofficial exchange rates in a fixed exchange rate system.
 - b. Of *buy, sell, or do nothing*, a central bank will likely do this with its reserve currency if excess demand for its own currency leads to illegal trades at a higher value.
 - c. Of *credible or not credible*, this describes a fixed exchange rate system that coexists with a black market.

Chapter 23

Policy Effects with Fixed Exchange Rates

Government policies work differently under a system of fixed exchange rates rather than floating rates. Monetary policy can lose its effectiveness whereas fiscal policy can become supereffective. In addition, fixed exchange rates offer another policy option, namely, exchange rate policy. Even though a fixed exchange rate should mean the country keeps the rate fixed, sometimes countries periodically change their fixed rate.

This chapter considers these policies under the assumptions of the AA-DD model. It concludes with a case study about the decline of the Bretton Woods fixed exchange rate system that was in place after World War II.

23.1 Overview of Policy with Fixed Exchange Rates

LEARNING OBJECTIVE

1. Preview the comparative statics results from the AA-DD model with fixed exchange rates.

This chapter uses the AA-DD model to describe the effects of fiscal, monetary, and exchange rate policy under a system of fixed exchange rates. Fiscal and monetary policies are the primary tools governments use to guide the macroeconomy. With fixed exchange rates, a third policy option becomes available—that is, exchange rate policy. Thus we also examine the effects of changes in the fixed exchange rate. These exchange rate changes are called **devaluations**¹ (sometimes competitive devaluations) and **revaluations**².

In introductory macroeconomics courses, students learn how government policy levers can be used to influence the level of the gross national product (GNP), inflation rate, unemployment rate, and interest rates. In this chapter, that analysis is expanded to an open economy (i.e., one open to trade) and to the effects on exchange rates and current account balances.

Results

Using the AA-DD model, several important relationships between key economic variables are shown:

- A monetary policy (change in M^S) has no effect on GNP or the exchange rate in a fixed exchange system. As such, the trade balance, unemployment, and interest rates all remain the same as well. Monetary policy becomes ineffective as a policy tool in a fixed exchange rate system.
- Expansionary fiscal policy ($\uparrow G$, $\uparrow TR$, or $\downarrow T$) causes an increase in GNP while maintaining the fixed exchange rate and constant interest rates. The trade balance and unemployment are both reduced.
- Contractionary fiscal policy ($\downarrow G$, $\downarrow TR$, or $\uparrow T$) reduces GNP while maintaining the fixed exchange rate and constant interest rates. The trade balance and unemployment both rise.

1. When the government lowers the value of its currency with respect to the reserve currency or to gold in a fixed exchange rate system.
2. When the government raises the value of its currency with respect to the reserve currency or to gold in a fixed exchange rate system.

- A competitive devaluation lowers the currency value and causes an increase in GNP. Unemployment falls, interest rates remain the same, and the trade balance rises.
- A currency revaluation raises the currency value and causes a decrease in GNP. Unemployment rises, interest rates remain the same, and the trade balance falls.
- Monetary expansion by the reserve currency country forces the domestic country to run a balance of payments surplus to maintain its fixed exchange rate. The resulting money supply increase causes domestic interest rates to fall to maintain equality with the falling foreign interest rates. Domestic GNP remains fixed, as do unemployment and the trade balance.
- A currency crisis arises when a country runs persistent balance of payments deficits while attempting to maintain its fixed exchange rate and is about to deplete its foreign exchange reserves. A crisis can force a country to devalue its currency or move to a floating exchange rate. To postpone the crisis, countries can sometimes borrow money from organizations like the International Monetary Fund (IMF).
- Anticipation of a **balance of payments crisis**³ can induce investors to sell domestic assets in favor of foreign assets. This is called **capital flight**⁴. Capital flight will worsen a balance of payments problem and can induce a crisis to occur.

Connections

The AA-DD model was developed to describe the interrelationships of macroeconomic variables within an open economy. Since some of these macroeconomic variables are controlled by the government, we can use the model to understand the likely effects of government policy changes. The main levers the government controls are monetary policy (changes in the money supply), fiscal policy (changes in the government budget), and exchange rate policy (setting the fixed exchange rate value). In this chapter, the AA-DD model is applied to understand government policy effects in the context of a fixed exchange rate system. In [Chapter 21 "Policy Effects with Floating Exchange Rates"](#), we considered these same government policies in the context of a floating exchange rate system. In [Chapter 24 "Fixed versus Floating Exchange Rates"](#), we'll compare fixed and floating exchange rate systems and discuss the pros and cons of each system.

3. Occurs when the country is about to run out of foreign exchange reserves because of persistent balance of payments deficits.

4. Refers to investors purchasing assets abroad in anticipation of an imminent currency devaluation or depreciation, often in the midst of a balance of payments crisis.

It is important to recognize that these results are what “would” happen under the full set of assumptions that describe the AA-DD model. These effects may or may not happen in reality. Nevertheless, the model surely captures some of the simple cause-and-effect relationships and therefore helps us to understand the broader implications of policy changes. Thus even if in reality many more elements (not

described in the model) may act to influence the key endogenous variables, the AA-DD model at least gives a more complete picture of some of the expected tendencies.

KEY TAKEAWAYS

- The main objective of the AA-DD model is to assess the effects of monetary, fiscal, and exchange rate policy changes.
- It is important to recognize that these results are what “would” happen under the full set of assumptions that describes the AA-DD model; they may or may not accurately describe actual outcomes in actual economies.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *appreciation, depreciation, or no change*, the effect of expansionary monetary policy on the domestic currency value under fixed exchange rates in the AA-DD model.
 - b. Of *increase, decrease, or no change*, the effect of contractionary monetary policy on GNP under fixed exchange rates in the AA-DD model.
 - c. Of *increase, decrease, or no change*, the effect of expansionary monetary policy on the current account deficit under fixed exchange rates in the AA-DD model.
 - d. Of *increase, decrease, or no change*, the effect of contractionary monetary policy on the current account surplus under fixed exchange rates in the AA-DD model.
 - e. Of *appreciation, depreciation, or no change*, the effect of expansionary fiscal policy on the domestic currency value under fixed exchange rates in the AA-DD model.
 - f. Of *increase, decrease, or no change*, the effect of contractionary fiscal policy on GNP under fixed exchange rates in the AA-DD model.
 - g. Of *increase, decrease, or no change*, the effect of expansionary fiscal policy on the current account deficit under fixed exchange rates in the AA-DD model.
 - h. Of *increase, decrease, or no change*, the effect of a devaluation on GNP under fixed exchange rates in the AA-DD model.
 - i. Of *increase, decrease, or no change*, the effect of a revaluation on the current account deficit under fixed exchange rates in the AA-DD model.
 - j. The term used to describe a rapid purchase of foreign investments often spurred by the expectation of an imminent currency devaluation.
 - k. The term used to describe the situation when a central bank runs persistent balance of payments deficits and is about to run out of foreign exchange reserves.

23.2 Monetary Policy with Fixed Exchange Rates

LEARNING OBJECTIVE

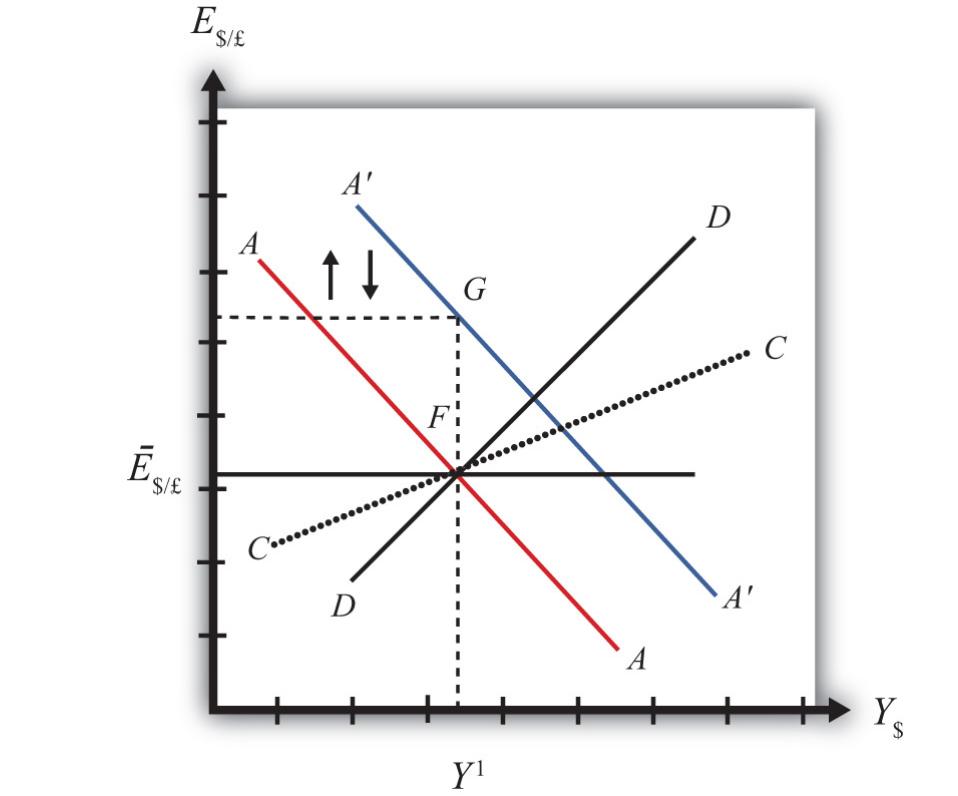
1. Learn how changes in monetary policy affect GNP, the value of the exchange rate, and the current account balance in a fixed exchange rate system in the context of the AA-DD model.
2. Understand the adjustment process in the money market, the Forex market, and the G&S market.

In this section, we use the AA-DD model to assess the effects of monetary policy in a fixed exchange rate system. Recall from [Chapter 18 "Interest Rate Determination"](#) that the money supply is effectively controlled by a country's central bank. In the case of the United States, this is the Federal Reserve Board, or the Fed. When the money supply increases due to action taken by the central bank, we refer to it as expansionary monetary policy. If the central bank acts to reduce the money supply, it is referred to as contractionary monetary policy. Methods that can be used to change the money supply are discussed in [Chapter 18 "Interest Rate Determination", Section 18.5 "Controlling the Money Supply"](#).

Expansionary Monetary Policy

Suppose the United States fixes its exchange rate to the British pound at the rate $\bar{E}_{\$/\text{£}}$. This is indicated in [Figure 23.1 "Expansionary Monetary Policy with a Fixed Exchange Rate"](#) as a horizontal line drawn at $\bar{E}_{\$/\text{£}}$. Suppose also that the economy is originally at a superequilibrium shown as point F with original gross national product (GNP) level Y^1 . Next, suppose the U.S. central bank (the Fed) decides to expand the money supply by conducting an open market operation, *ceteris paribus*. *Ceteris paribus* means that all other exogenous variables are assumed to remain at their original values. A purchase of Treasury bonds by the Fed will lead to an increase in the dollar money supply. As shown in [Chapter 20 "The AA-DD Model", Section 20.5 "Shifting the AA Curve"](#), money supply changes cause a shift in the AA curve. More specifically, an increase in the money supply will cause AA to shift upward (i.e., $\uparrow M^S$ is an AA up-shifter). This is depicted in the diagram as a shift from the red AA to the blue $A'A'$ line.

Figure 23.1 Expansionary Monetary Policy with a Fixed Exchange Rate



The money supply increase puts upward pressure on the exchange rate in the following way. First, a money supply increase causes a reduction in U.S. interest rates. This in turn reduces the rate of return on U.S. assets below the rate of return on similar assets in Britain. Thus international investors will begin to demand more pounds in exchange for dollars on the private Forex to take advantage of the relatively higher RoR of British assets. In a floating exchange system, excess demand for pounds would cause the pound to appreciate and the dollar to depreciate. In other words, the exchange rate $E_{\$/\text{£}}$ would rise. In the diagram, this would correspond to a movement to the new $A'A'$ curve at point G .

However, because the country maintains a fixed exchange rate, excess demand for pounds on the private Forex will automatically be relieved by Fed intervention. The Fed will supply the excess pounds demanded by selling reserves of pounds in exchange for dollars at the fixed exchange rate. As we showed in [Chapter 21 "Policy Effects with Floating Exchange Rates"](#), [Section 21.5 "Foreign Exchange Interventions with Floating Exchange Rates"](#), Fed sales of foreign currency result in a reduction in the U.S. money supply. This is because when the Fed buys dollars in the private Forex, it is taking those dollars out of circulation and thus out of the money supply. Since a reduction of the money supply causes AA to shift back down,

the final effect will be that the AA curve returns to its original position. This is shown as the up and down movement of the AA curve in the diagram. The final equilibrium is the same as the original at point F.

The AA curve must return to the same original position because the exchange rate must remain fixed at $\bar{E}_{\$/\text{£}}$. This implies that the money supply reduction due to Forex intervention will exactly offset the money supply expansion induced by the original open market operation. Thus the money supply will temporarily rise but then will fall back to its original level. Maintaining the money supply at the same level also assures that interest rate parity is maintained. Recall that in a fixed exchange rate system, interest rate parity requires equalization of interest rates between countries (i.e., $i_{\$} = i_{\text{£}}$). If the money supply did not return to the same level, interest rates would not be equalized.

Thus after final adjustment occurs, there are no effects from expansionary monetary policy in a fixed exchange rate system. The exchange rate will not change and there will be no effect on equilibrium GNP. Also, since the economy returns to the original equilibrium, there is also no effect on the current account balance.

Contractionary Monetary Policy

Contractionary monetary policy corresponds to a decrease in the money supply or a Fed sale of Treasury bonds on the open bond market. In the AA-DD model, a decrease in the money supply shifts the AA curve downward. The effects will be the opposite of those described above for expansionary monetary policy. A complete description is left for the reader as an exercise.

*The quick effects, however, are as follows. U.S. contractionary monetary policy with a fixed exchange rate will have **no effects** within the economy. $E_{\$/\text{£}}$, $Y_{\$}$ and the current account balance will all be maintained or return to their initial levels.*

Discussion

This result indicates that monetary policy is ineffective in influencing the economy in a fixed exchange rate system. In contrast, in a floating exchange rate system, monetary policy can either raise or lower GNP, at least in the short run. Thus monetary policy has some effectiveness in a floating system, and central bank authorities can adjust policy to affect macroeconomic conditions within their economy. For example, if the economy is growing only sluggishly, or perhaps is contracting, the central bank can raise the money supply to help spur an expansion of GNP, if the economy has a floating exchange rate. However, with a fixed

exchange rate, the central bank no longer has this ability. This explains why countries lose monetary autonomy (or independence) with a fixed exchange rate. The central bank can no longer have any influence over the interest rate, exchange rate, or the level of GNP.

One other important comparison worth making is between expansionary monetary policy in a fixed exchange rate system with sterilized foreign exchange (Forex) interventions in a floating system. In the first case, expansionary monetary policy is offset later with a contraction of the money supply caused by automatic Forex intervention. In the second case, Forex intervention leading to an expansion of the money supply is countered with contractionary open market operations. In the first case, the interest rate is maintained to satisfy interest rate parity. In the second case, the interest rate remains fixed by design. Clearly, these two situations represent exactly the same set of actions, though in a different order. Thus it makes sense that the two policies would have the same implications—that is, “no impact” on any of the economic variables.

KEY TAKEAWAYS

- There are *no effects* from expansionary or contractionary monetary policy in a fixed exchange rate system. The exchange rate will not change, there will be no effect on equilibrium GNP, and there will be no effect on the current account balance.
- Monetary policy in a fixed exchange rate system is equivalent in its effects to sterilized Forex interventions in a floating exchange rate system.

EXERCISE

1. Suppose that Latvia can be described with the AA-DD model and that Latvia fixes its currency, the lats (Ls), to the euro. Consider the changes in the exogenous variable in the left column. Indicate the short-run effects on the equilibrium levels of Latvian GNP, the Latvian interest rate (i_{Ls}), the Latvian trade balance, and the exchange rate (E_{Ls}/ϵ). Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	GNP	i_{Ls}	Trade Balance	E_{Ls}/ϵ
An increase in the Latvian money supply				
A decrease in the Latvian money supply				

23.3 Fiscal Policy with Fixed Exchange Rates

LEARNING OBJECTIVES

1. Learn how changes in fiscal policy affect GNP, the value of the exchange rate, and the current account balance in a fixed exchange rate system in the context of the AA-DD model.
2. Understand the adjustment process in the money market, the Forex market, and the G&S market.

In this section, we use the AA-DD model to assess the effects of fiscal policy in a fixed exchange rate system. Recall from [Chapter 19 "National Output Determination"](#) that fiscal policy refers to any change in expenditures or revenues within any branch of the government. This means any change in government spending (e.g., transfer payments or taxes) by federal, state, or local governments represents a fiscal policy change. Since changes in expenditures or revenues will often affect a government budget balance, we can also say that a change in the government surplus or deficit represents a change in fiscal policy.

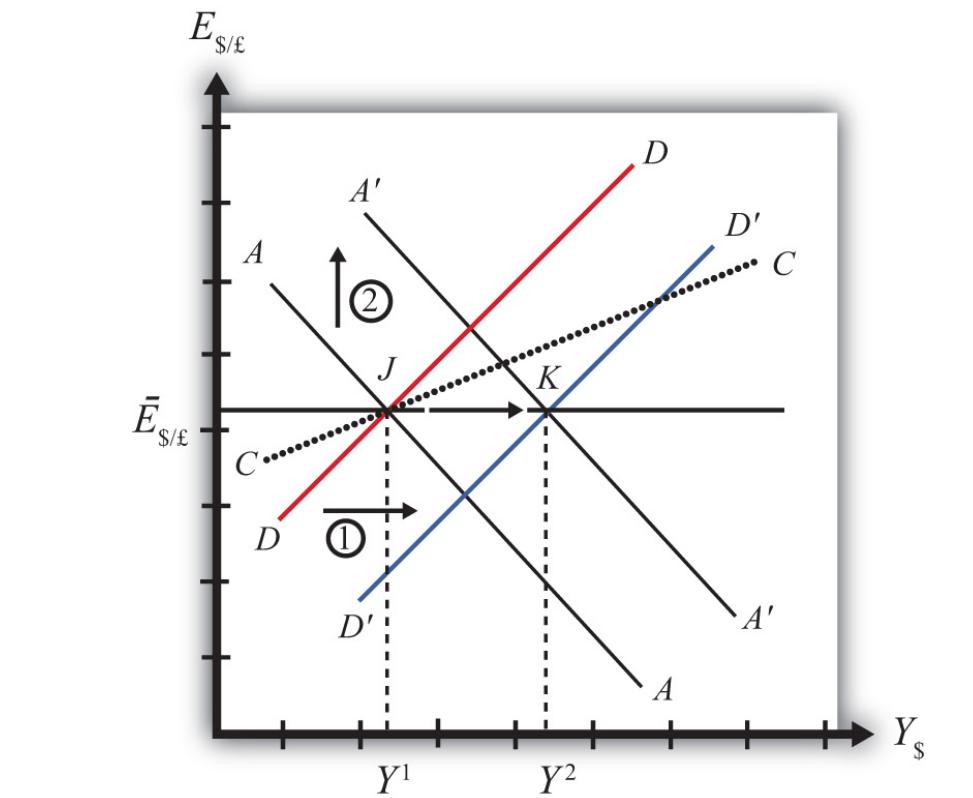
When government spending or transfer payments increase, or tax revenues decrease, we refer to it as expansionary fiscal policy. These actions would also be associated with an increase in the government budget deficit, or a decrease in its budget surplus. If the government acts to reduce government spending or transfer payments, or increase tax revenues, it is referred to as contractionary fiscal policy. These actions would also be associated with a decrease in the government budget deficit, or an increase in its budget surplus.

Expansionary Fiscal Policy

Suppose the United States fixes its exchange rate to the British pound at the rate \bar{E}/ϵ . This is indicated in [Figure 23.2 "Expansionary Fiscal Policy with a Fixed Exchange Rate"](#) as a horizontal line drawn at \bar{E}/ϵ . Suppose also that the economy is originally at a superequilibrium shown as point J with GNP at level Y^1 . Next, suppose the government decides to increase government spending (or increase transfer payments or decrease taxes). As shown in [Chapter 20 "The AA-DD Model", Section 20.3 "Shifting the DD Curve"](#), fiscal policy changes cause a shift in the DD curve. More specifically, an increase in government spending (or an increase in transfer payments or a decrease in taxes) will cause DD to shift rightward (i.e., $\uparrow G$,

$\uparrow TR$, and $\downarrow T$ all are DD right-shifters). This is depicted in the diagram as a shift from the red DD to the blue $D'D'$ line (step 1).

Figure 23.2 Expansionary Fiscal Policy with a Fixed Exchange Rate



If the expansionary fiscal policy occurs because of an increase in government spending, then government demand for goods and services (G&S) will increase. If the expansionary fiscal policy occurs due to an increase in transfer payments or a decrease in taxes, then disposable income will increase, leading to an increase in consumption demand. In either case, aggregate demand increases. Before any adjustment occurs, the increase in aggregate demand causes aggregate demand to exceed aggregate supply, which will lead to an expansion of GNP. Thus the economy will begin to move rightward from point J.

As GNP rises, so does real money demand, causing an increase in U.S. interest rates. With higher interest rates, the rate of return on U.S. assets rises above that in the United Kingdom and international investors increase demand for dollars (in exchange for pounds) on the private Forex. In a floating exchange rate system this would lead to a U.S. dollar appreciation (and pound depreciation)—that is, a decrease in the exchange rate $E_{\$/£}$.

However, because the country maintains a fixed exchange rate, excess demand for dollars on the private Forex will automatically be relieved by the U.S. Federal Reserve (or the Fed) intervention. The Fed will supply the excess dollars demanded by buying pounds in exchange for dollars at the fixed exchange rate. As we showed in [Chapter 21 "Policy Effects with Floating Exchange Rates"](#), [Section 21.5 "Foreign Exchange Interventions with Floating Exchange Rates"](#), the foreign currency purchases by the Fed result in an increase in the U.S. money supply. This is because when the Fed sells dollars in the private Forex, these dollars are entering into circulation and thus become a part of the money supply. The increase in the money supply causes the AA curve to shift up (step 2). The final equilibrium will be reached when the new A'A' curve intersects the D'D' curve at the fixed exchange rate ($\bar{E}_{\$/\text{£}}$) shown at point K.

Note that in the transition, the Fed intervention in the Forex occurred because investors responded to rising U.S. interest rates by increasing demand for dollars on the Forex. The Fed's response causes an increase in the money supply, which in turn will lower interest rates back to their original level. This result is necessary to maintain the fixed exchange rate interest rate parity (IRP) condition of $i_{\$} = i_{\text{£}}$.

Note also that as GNP increases in the transition, causing interest rates to rise, this rise is immediately countered with automatic Fed intervention in the Forex. Thus the exchange rate will never fall below the fixed rate. There will be pressure for the exchange rate to fall, but the Fed will always be there to relieve the pressure with its intervention. Thus the adjustment path from the original equilibrium at J to the final equilibrium at K will follow the rightward arrow between the two points along the fixed exchange rate.

The final result is that *expansionary fiscal policy in a fixed exchange rate system will cause an increase in GNP (from Y^1 to Y^2) and no change in the exchange rate in the short run*. Since the new equilibrium at K lies below the original CC curve representing a fixed current account balance, *expansionary fiscal policy*, consisting of an increase in G , *will cause the current account balance to fall*. This corresponds to a decrease in a trade surplus or an increase in a trade deficit.

Contractionary Fiscal Policy

Contractionary fiscal policy corresponds to a decrease in government spending, a decrease in transfer payments, or an increase in taxes. It would also be represented by a decrease in the government budget deficit or an increase in the budget surplus. In the AA-DD model, a contractionary fiscal policy shifts the DD curve leftward. The effects will be the opposite of those described above for expansionary fiscal policy. A complete description is left for the reader as an exercise.

The quick effects, however, are as follows. *Contractionary fiscal policy in a fixed exchange rate system will cause a decrease in GNP and no change in the exchange rate in the short run.* Contractionary fiscal policy, consisting of a decrease in G , will also cause the current account balance to rise. This corresponds to an increase in a trade surplus or a decrease in a trade deficit.

KEY TAKEAWAYS

- Expansionary fiscal policy in a fixed exchange rate system will cause an increase in GNP, no change in the exchange rate (of course), and a decrease in the current account balance.
- Contractionary fiscal policy in a fixed exchange rate system will cause a decrease in GNP, no change in the exchange rate (of course), and an increase in the current account balance.

EXERCISES

1. Sri Lanka fixes its currency, the Sri Lankan rupee (LKR), to the U.S. dollar. Suppose Sri Lanka can be described using the AA-DD model. Consider changes in the exogenous variables in Sri Lanka in the left column. Suppose each change occurs *ceteris paribus*. Indicate the short-run effects on the equilibrium values of Sri Lankan GNP, the Sri Lankan interest rate (i_{LKR}), the Sri Lankan trade deficit, and the exchange rate ($E_{LKR}/\$$). Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	GNP	i_{LKR}	Sri Lankan Trade Deficit	$E_{LKR}/\$$
A decrease in domestic taxes				
An increase in government demand				
An increase in transfer payments				

2. Consider the following occurrences. Use the AA-DD model to determine the impact on the variables (+, -, 0, or A) from the twin-deficit identity listed along the top row. Consider only short-run effects (i.e., before inflationary effects occur) and assume *ceteris paribus* for all other exogenous variables.

	Impact on			
	S_p	I	$IM - EX$	$G + TR - T$
A reduction in government spending with a fixed exchange rate				
An increase in transfer payments with fixed exchange rates				
A decrease in taxes with fixed exchange rates				

23.4 Exchange Rate Policy with Fixed Exchange Rates

LEARNING OBJECTIVES

1. Learn how changes in exchange rate policy affect GNP, the value of the exchange rate, and the current account balance in a fixed exchange rate system in the context of the AA-DD model.
2. Understand the adjustment process in the money market, the Forex market, and the G&S market.

In this section, we use the AA-DD model to assess the effects of exchange rate policy in a fixed exchange rate system. In a sense we can say that the government's decision to maintain a fixed exchange *is* the country's exchange rate policy. However, over time, the government does have some discretion concerning the value of the exchange rate. In this section, we will use "exchange rate policy" to mean changes in the value of the fixed exchange rate.

If the government lowers the value of its currency with respect to the reserve currency or to gold, we call the change a devaluation. If the government raises the value of its currency with respect to the reserve currency or to gold, we call the change a revaluation. The terms *devaluation* and *revaluation* should properly be used only in reference to a government change in the fixed exchange rate since each term suggests an action being taken. In contrast, natural market changes in supply and demand will result in changes in the exchange rate in a floating system, but it is not quite right to call these changes devaluations or revaluations since no concerted action was taken by anyone. Nonetheless, some writers will sometimes use the terms this way.

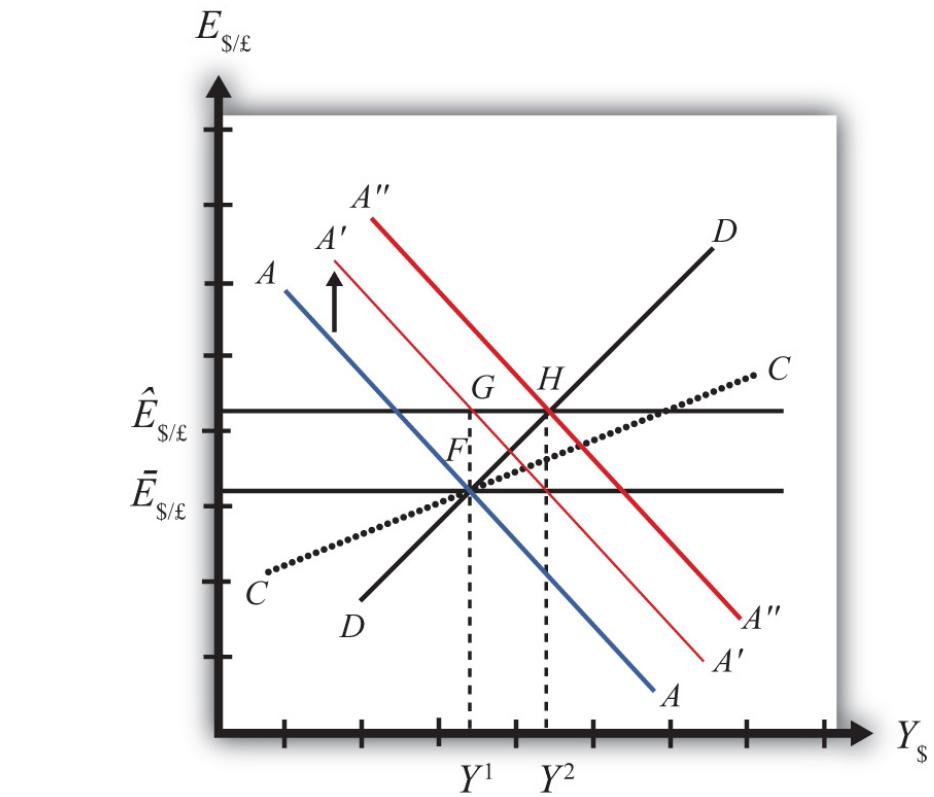
In most cases, devaluations and revaluations occur because of persistent balance of payments disequilibria. We will consider these situations in [Chapter 23 "Policy Effects with Fixed Exchange Rates"](#), [Section 23.6 "Currency Crises and Capital Flight"](#) on balance of payments crises and capital flight. In this section, we will consider the basic effects of devaluations and revaluations without assuming any notable prior events caused these actions to occur.

Devaluation

Suppose the United States fixes its exchange rate to the British pound at the rate \bar{E}/ϵ . This is indicated in [Figure 23.3 "Effects of a Devaluation"](#) as a horizontal line

drawn at $\bar{E}_{\$/\text{£}}$. Suppose also that the economy is originally at a superequilibrium shown as point F with gross national product (GNP) at level Y^1 . Next, suppose the U.S. central bank (or the Fed) decides to devalue the U.S. dollar with respect to the British pound corresponding to an increase in the fixed rate from $\bar{E}_{\$/\text{£}}$ to $\hat{E}_{\$/\text{£}}$. Recall that a devaluation corresponds to an increase in the $\$/\text{£}$ exchange rate. Assume that there was no anticipation of the devaluation and that it comes about as a complete surprise to all market participants.

Figure 23.3 Effects of a Devaluation



The first effect of the devaluation, of course, is that the exchange rate rises. Immediately the economy moves from F to G on the diagram. It may seem that this would move the economy off the AA curve, but instead the AA curve shifts up with the devaluation to $A'A'$. This occurs because the AA curve is a function of the expected exchange rate. As long as investors believe that the new exchange rate will now remain fixed at its new rate ($\hat{E}_{\$/\text{£}}$), the expected future exchange rate will immediately rise to this new level as well. It is this increase in $E_{\$/\text{£}}^e$ that causes AA to shift up.

When at point *G*, however, the economy is not at a superequilibrium. Because of the dollar devaluation, the real exchange rate has increased, making foreign goods relatively more expensive and U.S. goods relatively cheaper. This raises aggregate demand, which at the new exchange rate ($\hat{E}_{\$/\text{£}}$) is now at the level where the exchange rate line crosses the DD curve at point *H*.

Since the economy, for now, lies at *G* to the left of point *H* on the DD curve, aggregate demand exceeds supply. Producers will respond by increasing supply to satisfy the demand, and GNP will begin to rise.

As GNP rises, real money demand will rise, causing an increase in U.S. interest rates, which will raise the rate of return on U.S. assets. Investors will respond by increasing their demand for U.S. dollars on the foreign exchange (Forex) market, and there will be pressure for a dollar appreciation.

To maintain the fixed exchange rate, however, the U.S. Fed will have to automatically intervene on the Forex and sell dollars to satisfy the excess demand in exchange for pounds. This represents a balance of payments surplus since by buying pounds on the Forex the United States is adding to its stock of foreign reserves. A balance of payments surplus in turn causes an increase in the U.S. money supply, which will shift the AA curve to the right.

As GNP rises toward Y^2 at point *H*, the AA curve will shift right with the Fed intervention to maintain the equilibrium exchange rate at the new fixed value, which is $\hat{E}_{\$/\text{£}}$. The final superequilibrium occurs at point *H* where excess aggregate demand is finally satisfied.

The final result is that *a devaluation in a fixed exchange rate system will cause an increase in GNP (from Y^1 to Y^2) and an increase in the exchange rate to the new fixed value in the short run*. Since the new equilibrium at *H* lies above the original CC curve representing a fixed current account balance, *a devaluation will cause the current account balance to rise*. This corresponds to an increase in a trade surplus or a decrease in a trade deficit.

Revaluation

A revaluation corresponds to change in the fixed exchange rate such that the country's currency value is increased with respect to the reserve currency. In the AA-DD model, a U.S. dollar revaluation would be represented as a decrease in the fixed $\$/\text{£}$ exchange rate. The effects will be the opposite of those described above for a devaluation. A complete description is left for the reader as an exercise.

The quick effects, however, are as follows. A *revaluation in a fixed exchange rate system will cause a decrease in GNP and a decrease in the fixed exchange rate in the short run*. A revaluation will also cause the current account balance to fall. This corresponds to a decrease in a trade surplus or an increase in a trade deficit.

KEY TAKEAWAYS

- If the government lowers (raises) the value of its currency with respect to the reserve currency, or to gold, we call the change a devaluation (revaluation).
- A devaluation in a fixed exchange rate system will cause an increase in GNP, an increase in the exchange rate to the new fixed value in the short run, and an increase in the current account balance.
- A revaluation in a fixed exchange rate system will cause a decrease in GNP, an increase in the currency value to the new fixed rate, and a decrease in the current account balance.

EXERCISES

1. Vietnam fixes its currency, the Vietnamese dong (VND), to the US dollar. Suppose Vietnam can be described using the AA-DD model. Consider changes in the exogenous variables in Vietnam in the left column. Suppose each change occurs *ceteris paribus*. Indicate the short-run effects on the equilibrium values of Vietnamese GNP, the Vietnamese interest rate (i_{VND}), the Vietnamese trade deficit, and the exchange rate ($E_{VND}/\$/$). Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	GNP	i_{VND}	$E_{VND}/\$$
A devaluation of the Vietnamese dong			
A revaluation of the Vietnamese dong			

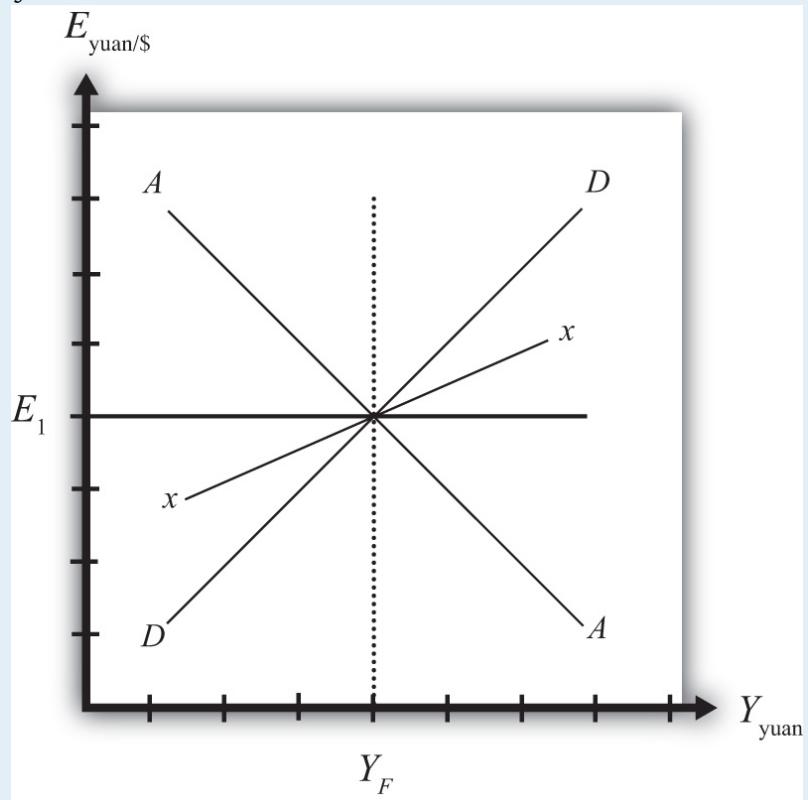
2. Consider the following occurrences. Use the AA-DD model to determine the impact on the variables (+, -, 0, or A) from the twin-deficit identity listed along the top row. Consider only short-run effects (i.e., before inflationary effects occur) and assume *ceteris paribus* for all other exogenous variables.

	Impact on			
	S_p	I	$IM - EX$	$G + TR - T$
A currency devaluation under fixed exchange rates				

	Impact on			
	S_p	I	$IM - EX$	$G + TR - T$
A currency revaluation under fixed exchange rates				

3. China maintains an exchange rate fixed to the U.S. dollar at the rate E_1 . Use the following AA-DD diagram for China to depict answers to the questions below. Suppose China's current account is in surplus originally. Suppose Y_F indicates the full employment level of output.

Figure 23.4



- a. Suppose China unexpectedly *revalues* its currency under pressure from the U.S. government. Draw a line for the new exchange rate and mark the graph with an E_2 .

- b. Mark the graph with a T to indicate the position of the economy immediately after the revaluation when investor expectations adjust to the new exchange rate.
- c. What effect does the revaluation have for the prices of Chinese goods to Americans?
- d. Mark the graph with a W to indicate the position of the economy once a new short-run equilibrium is achieved. Mark the graph with Y_2 to indicate the new level of GDP.
- e. Does China's stock of foreign reserves *rise* or *fall* after the revaluation?
- f. Does China's current account surplus *rise* or *fall*?
- g. In the adjustment to a long-run equilibrium, would the Chinese price level *rise* or *fall*?

23.5 Reserve Country Monetary Policy under Fixed Exchange Rates

LEARNING OBJECTIVES

1. Learn how monetary policy in the foreign reserve country affects domestic GNP, the value of the exchange rate, and the current account balance in a fixed exchange rate system in the context of the AA-DD model.
2. Understand the adjustment process in the money market, the Forex market, and the G&S market.

Suppose the United States fixes its exchange rate to the British pound. In this circumstance, the exchange rate system is a reserve currency standard in which the British pound is the reserve currency. The U.S. government is the one that fixes its exchange rate and will hold some quantity of British pounds on reserve so it is able to intervene on the Forex to maintain the credible fixed exchange rate.

It is worth noting that since the United States fixes its exchange rate to the pound, the British pound is, of course, fixed to the U.S. dollar as well. Since the pound is the reserve currency, however, it has a special place in the monetary system. The Bank of England, Britain's central bank, will never need to intervene in the Forex market. It does not need to hold dollars. Instead, all market pressures for the exchange rate to change will be resolved by U.S. intervention, that is, by the nonreserve currency country.

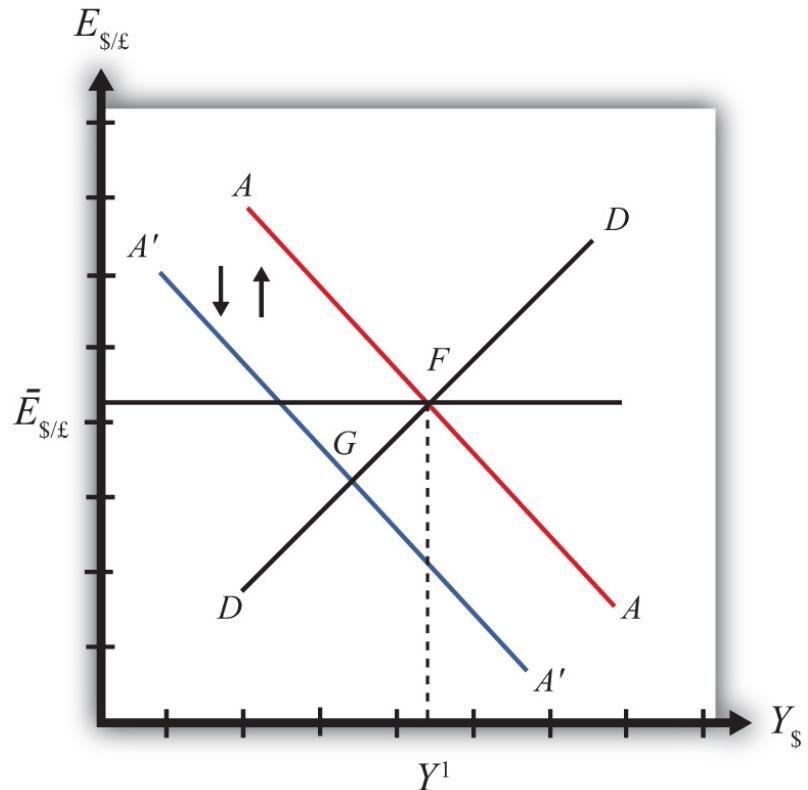
Expansionary Monetary Policy by the Reserve Country

Now let's suppose that the reserve currency country, Britain, undertakes expansionary monetary policy. We will consider the impact of this change from the vantage point of the United States, the nonreserve currency country. Suppose the United States is originally in a superequilibrium at point F in the adjoining diagram with the exchange rate fixed at $\bar{E}_{\$/\text{£}}$. An increase in the British money supply will cause a decrease in British interest rates, $i_{\text{£}}$.

As shown in [Chapter 20 "The AA-DD Model"](#), [Section 20.5 "Shifting the AA Curve"](#), foreign interest rate changes cause a shift in the AA curve. More specifically, a decrease in the foreign interest rate will cause the AA curve to shift downward (i.e.,

$\downarrow i_E$ is an AA down-shifter). This is depicted in [Figure 23.5 "Expansionary Monetary Policy by a Reserve Country"](#) as a shift from the red AA to the blue A'A' line.

Figure 23.5 Expansionary Monetary Policy by a Reserve Country



The money supply decrease puts downward pressure on the exchange rate in the following way. When British interest rates fall, it will cause $i_E < i_S$ and interest rate parity (IRP) will be violated. Thus international investors will begin to demand more dollars in exchange for pounds on the private Forex to take advantage of the relatively higher rate of return on U.S. assets. In a floating exchange system, excess demand for dollars would cause the dollar to appreciate and the pound to depreciate. In other words, the exchange rate ($E_{\$/\text{£}}$) would fall. In the diagram, this would correspond to a movement to the new A'A' curve at point G.

Because the country maintains a fixed exchange rate, however, excess demand for dollars on the private Forex will automatically be relieved by the U.S. Federal Reserve (or the Fed) intervention. The Fed will supply the excess dollars demanded by buying pounds in exchange for dollars at the fixed exchange rate. As we showed in [Chapter 21 "Policy Effects with Floating Exchange Rates"](#), [Section 21.5 "Foreign Exchange Interventions with Floating Exchange Rates"](#), the foreign currency

purchases by the Fed result in an increase in the U.S. money supply. This is because when the Fed sells dollars in the private Forex, these dollars are entering into circulation and thus become a part of the money supply. Since an increase in the money supply causes AA to shift up, the AA curve will return to its original position to maintain the fixed exchange rate. This is shown as the up-and-down movement of the AA curve in the diagram. Thus the final equilibrium is the same as the original equilibrium at point F.

Remember that in a fixed exchange rate system, IRP requires equalization of interest rates between countries. When the British interest rates fell, they fell below the rates in the United States. When the U.S. Fed intervenes on the Forex, however, the U.S. money supply rises and U.S. interest rates are pushed down. Pressure for the exchange rate to change will cease only when U.S. interest rates become equal to British interest rates and IRP ($i_E = i_{\$}$) is again satisfied.

Thus after final adjustment occurs, expansionary monetary policy by the foreign reserve currency country in a fixed exchange rate system causes no effects on U.S. GNP or the exchange rate. Since the economy also returns to the original equilibrium, there is also no effect on the current account balance. Fed intervention in the Forex to maintain the fixed exchange rate, however, will cause U.S. interest rates to fall to maintain IRP with the lower reserve country interest rates.

Contractionary Monetary Policy by the Reserve Country

Contractionary monetary policy corresponds to a decrease in the British money supply that would lead to an increase in British interest rates. In the AA-DD model, an increase in foreign interest rates shifts the AA curve upward. The effects will be the opposite of those described above for expansionary monetary policy. A complete description is left for the reader as an exercise.

KEY TAKEAWAYS

- Expansionary monetary policy by the foreign reserve currency country in a fixed exchange rate system causes no effects on domestic GNP, the exchange rate, or the current account balance in the AA-DD model. However, it will cause domestic interest rates to fall.
- Contractionary monetary policy by the foreign reserve currency country in a fixed exchange rate system causes no effects on domestic GNP, the exchange rate, or the current account balance in the AA-DD model. However, it will cause domestic interest rates to rise.

EXERCISES

1. Honduras fixes its currency, the Honduran lempira (HNL), to the U.S. dollar. Suppose Honduras can be described using the AA-DD model. Consider changes in the exogenous variables in the left column. Suppose each change occurs *ceteris paribus*. Indicate the short-run effects on the equilibrium values of Honduran GNP, the Honduran interest rate (i_{HNL}), the Honduran trade deficit, and the exchange rate ($E_{\text{HNL}}/\$$). Use the following notation:

+ the variable increases

- the variable decreases

0 the variable does not change

A the variable change is ambiguous (i.e., it may rise, it may fall)

	GNP	i_{HNL}	$E_{\text{HNL}}/\$$
An increase in U.S. interest rates			
A decrease in U.S. interest rates			

2. Consider the following occurrences. Use the AA-DD model to determine the impact on the variables (+, -, 0, or A) from the twin-deficit identity listed along the top row. Consider only short-run effects (i.e., before inflationary effects occur) and assume *ceteris paribus* for all other exogenous variables.

	Impact on			
	S_p	I	$IM - EX$	$G + TR - T$
An increase in foreign interest rates under fixed exchange rates				

	Impact on			
	S_p	I	$IM - EX$	$G + TR - T$
A decrease in foreign interest rates under fixed exchange rates				

23.6 Currency Crises and Capital Flight

LEARNING OBJECTIVE

1. Learn how currency crises develop and lead to capital flight.

To maintain a credible fixed exchange rate system, a country will need to buy and sell the reserve currency whenever there is excess demand or supply in the private foreign exchange (Forex). To make sales of foreign currency possible, a country will need to maintain a foreign exchange reserve. The reserve is a stockpile of assets denominated in the reserve currency. For example, if the United States fixes the dollar to the British pound, then it would need to have a reserve of pound assets in case it needs to intervene on the Forex with a sale of pounds.

Generally, a central bank holds these reserves in the form of Treasury bonds issued by the reserve country government. In this way, the reserve holdings earn interest for the central bank and thus the reserves will grow in value over time. Holding reserves in the form of currency would not earn interest and thus are less desirable. Nonetheless, a central bank will likely keep some of its reserves liquid in the form of currency to make anticipated daily Forex transactions. If larger sales of reserves become necessary, the U.S. central bank can always sell the foreign Treasury bonds on the bond market and convert those holdings to currency.

A fixed exchange rate is sustainable if the country's central bank can maintain that rate over time with only modest interventions in the Forex. Ideally, one would expect that during some periods of time, there would be excess demand for domestic currency on the Forex, putting pressure on the currency to appreciate. In this case, the central bank would relieve the pressure by selling domestic currency and buying the reserve on the Forex, thus running a balance of payments (BoP) surplus. During these periods, the country's reserve holdings would rise. At other periods, there may be excess demand for the reserve currency, putting pressure on the domestic currency to depreciate. Here, the central bank would relieve the pressure by selling the reserve currency in exchange for domestic currency, thus running a balance of payments deficit. During these periods, the country's reserve holdings would fall. As long as the country's reserve holdings stay sufficiently high during its ups and downs, the fixed exchange rate could be maintained indefinitely. In this way, the central bank's interventions "smooth-out" the fluctuations that would have occurred in a floating system.

Problems arise if the reserves cannot be maintained if, for example, there is a persistent excess demand for the foreign currency over time with very few episodes of excess supply. In this case, the central bank's persistent BoP deficits will move reserve holdings closer and closer to zero. A balance of payments crisis occurs when the country is about to run out of foreign exchange reserves.

Borrowing Reserves

Several things may happen leading up to a balance of payments crisis. One option open to the central bank is to borrow additional quantities of the reserve currency from the reserve country central bank, government, or an international institution like the International Monetary Fund (IMF). The IMF was originally created to help countries with balance of payments problems within the Bretton Woods fixed exchange rate system (1945–1973). When a country was near to depleting its reserves, it could borrow reserve currency from the IMF. As long as the balance of payments deficits leading to reserve depletion would soon be reversed with balance of payments surpluses, the country would be able to repay the loans to the IMF in the near future. As such, the IMF “window” was intended to provide a safety valve in case volatility in supply and demand in the Forex was greater than a country's reserve holdings could handle.

Devaluation

If a country cannot acquire additional reserves and if it does not change domestic policies in a way that causes excess demand for foreign currency to cease or reverse, then the country will run out of foreign reserves and will no longer be able to maintain a credible fixed exchange rate. The country could keep the fixed exchange rate at the same level and simply cease intervening in the Forex; however, this would not relieve the pressure for the currency to depreciate and would quickly create conditions for a thriving black market.

If the country remains committed to a fixed exchange rate system, its only choice is to devalue its currency with respect to the reserve. A lower currency value will achieve two things. First, it will reduce the prices of all domestic goods from the viewpoint of foreigners. In essence, a devaluation is like having a sale in which all the country's goods are marked down by some percentage. At the same time, the devaluation will raise the price of foreign goods to domestic residents. Thus foreign goods have all been marked up in price by some percentage. These changes should result in an increase in demand for domestic currency to take advantage of the lower domestic prices and a decrease in demand for foreign currency due to the higher foreign prices.

The second effect occurs for investors. When the currency is devalued, the rate of return on foreign assets may fall, especially if investors had anticipated a devaluation and had adjusted their expectations accordingly. (See the next section on capital flight for further discussion.) When the rate of return on foreign assets falls, the demand for foreign currency will also fall.

If the devaluation is large enough to reverse the currency demand in the Forex, generating excess demand for the domestic currency, the central bank will have to buy foreign reserves to maintain the new devalued exchange rate and can begin to accumulate a stockpile of reserves once again.

Capital Flight

Balance of payments crises are often anticipated by investors in the marketplace. When this occurs it will result in capital flight, which in turn is likely to aggravate the balance of payments crisis. Here's why.

The interest rate parity condition holds when rates of return on domestic and foreign assets are equalized. Recall from [Chapter 22 "Fixed Exchange Rates"](#), [Section 22.3 "Interest Rate Parity with Fixed Exchange Rates"](#) that in a fixed exchange rate system the IRP condition simplifies to equalization of interest rates between two countries. However, this result assumed that investors expected the currency to remain fixed indefinitely. If investors believe instead that a country is about to suffer a balance of payments crisis and run out of foreign reserves, they will also anticipate that a devaluation will occur soon.

Assume as before that the United States fixes its currency to the British pound. The interest rate parity condition can be written as

$$i_{\$} = i_{\text{£}} + (1 + i_{\text{£}}) \frac{E_{\$/\text{£}}^e - \bar{E}_{\$/\text{£}}}{\bar{E}_{\$/\text{£}}} ,$$

where the left side is the rate of return on U.S. assets, equal to the average U.S. interest rate, and the right side is the rate of return on British assets. When there is no imminent balance of payments crisis, investors should expect the future exchange rate ($E_{\$/\text{£}}^e$) to equal the current fixed exchange rate ($E_{\$/\text{£}}$) and the interest parity condition simplifies to $i_{\$} = i_{\text{£}}$. However, if investors recognize that the central bank is selling large quantities of its foreign reserves in the Forex regularly, then they are likely also to recognize that the balance of payments deficits are unsustainable. Once the reserves run out, the central bank will be forced to devalue its currency. Thus forward-looking investors should plan for that event.

today. The result is an increase in the expected exchange rate, above the current fixed rate, reflecting the expectation that the dollar will be devalued soon.

This, in turn, will increase the expected rate of return of British assets, raising the right side of the above expression. Now, $RoR_E > RoR_\$,$ and investors will increase demand for British pounds on the Forex. In this instance, investors are “fleeing” their own domestic assets to purchase foreign assets (or capital) that now have a greater expected return. Thus the action is called **capital flight**⁵.

The intuition for capital flight is simple. If an investor expects the domestic currency (and assets denominated in that currency) will soon fall in value, it is better to sell now before the value actually does fall. Also, as the domestic currency falls in value, the British pound is expected to rise in value. Thus it is wise to buy British pounds and assets while their prices are lower and profit on the increase in the pound value when the dollar devaluation occurs.

The broader effect of capital flight, which occurs in anticipation of a balance of payments crisis, is that it can actually force a crisis to occur much sooner. Suppose the United States was indeed running low on foreign reserves after running successive balance of payments deficits. Once investors surmise that a crisis may be possible soon and react with a change in their expected exchange rate, there will be a resulting increase in demand for pounds on the Forex. This will force the central bank to intervene even further in the Forex by selling foreign pound reserves to satisfy investor demand and to keep the exchange rate fixed. However, additional interventions imply an even faster depletion of foreign reserve holdings, bringing the date of crisis closer in time.

It is even possible for investor behavior to create a balance of payments crisis when one might not have occurred otherwise. Suppose the U.S. central bank (or the Fed) depletes reserves by running balance of payments deficits. However, suppose the Fed believes the reserve holdings remain adequate to defend the currency value, whereas investors believe the reserve holdings are inadequate. In this case, capital flight will likely occur that would deplete reserves much faster than before. If the capital flight is large enough, even if it is completely unwarranted based on market conditions, it could nonetheless deplete the remaining reserves and force the central bank to devalue the currency.

5. Refers to investors purchasing assets abroad in anticipation of an imminent currency devaluation or depreciation, often in the midst of a balance of payments crisis.

Return to Float

There is one other possible response for a country suffering from a balance of payments crisis. The country could always give up on the fixed exchange rate system and allow its currency to float freely. This means the central bank no longer

needs to intervene on the Forex and the exchange rate value will be determined by daily supply and demand conditions on the private Forex. Since the reason for the BoP crisis was continual pressure for the currency to depreciate, moving to a floating system would undoubtedly result in a rapidly depreciating currency.

The main advantage of returning to a floating exchange rate is that the private Forex market will quickly move the exchange rate to the level that equalizes supply and demand. In contrast, many times countries that devalue their fixed exchange rate do not devalue sufficiently and a second devaluation becomes necessary shortly thereafter. When the countries in the Bretton Woods system switched to floating rates in 1973, the original intention was to allow markets to adjust to the equilibrium exchange rates reflecting market conditions and then to refix the exchange rates at the sustainable equilibrium level. However, an agreement to reestablish fixed rates was never implemented. The U.S. dollar and many other currencies have been floating ever since.

A second advantage of switching to a floating system is that it relieves the central bank from the necessity of maintaining a stockpile of reserves. Thus the whole problem of balance of payments crises disappears completely once a country lets its currency float.

KEY TAKEAWAYS

- A fixed exchange rate is sustainable if the country's central bank can maintain that rate over time with only modest interventions in the Forex.
- A balance of payments crisis occurs when persistent balance of payments deficits bring a country close to running out of foreign exchange reserves.
- BoP crises can be resolved by (a) borrowing foreign reserves, (b) devaluation of the currency, or (c) moving to a floating exchange rate.
- In the midst of a BoP crisis, investors often purchase assets abroad in anticipation of an imminent currency devaluation or depreciation. This is known as capital flight.
- Capital flight works to exacerbate the BoP crisis because it results in a more rapid depletion of foreign exchange reserves and makes the crisis more likely to occur.

EXERCISES

1. List the three ways in which a balance of payments crisis can be resolved either temporarily or permanently. Which of these methods will be most effective, especially if the country continues to pursue the policies that led to the crisis?
2. Explain why capital flight, spurred by the expectation of a currency devaluation, can be a self-fulfilling prophecy.
3. If an expected currency devaluation inspires capital flight, explain what might happen if investors expect a currency revaluation.

23.7 Case Study: The Breakup of the Bretton Woods System, 1973

LEARNING OBJECTIVES

1. Learn how the Bretton Woods system of fixed exchange rates set up after World War II was supposed to work.
2. Learn how and why the system collapsed in 1973.
3. Recognize some of the problems inherent in one type of fixed exchange rate system.

In July 1944, delegates from forty-five of the allied powers engaged in World War II met in Bretton Woods, New Hampshire, in the United States to plan for the economic institutions believed necessary to assist in the reconstruction, development, and growth of the postwar economy. Foremost on the delegates' minds was the instability of the international economic system after World War I, including the experiences of hyperinflation as in Germany in 1922–1923 and the worldwide depression of the 1930s. One element believed necessary to avoid repeating the mistakes of the past was to implement a system of fixed exchange rates. Not only could fixed exchange rates help prevent inflation, but they could also eliminate uncertainties in international transactions and thus serve to promote the expansion of international trade and investment. It was further hoped that economic interconnectedness would make it more difficult for nationalism to reassert itself.

The *Bretton Woods system of exchange rates* was set up as a gold exchange standard, a cross between a pure gold standard and a reserve currency standard. In a gold exchange standard, one country is singled out to be the reserve currency. In the Bretton Woods case, the currency was the U.S. dollar. The U.S. dollar was fixed to a weight in gold, originally set at \$35 per ounce. The U.S. central bank agreed to exchange dollars for gold on demand, but only with foreign central banks. In a pure gold standard, the central bank would exchange gold for dollars with the general public as well.

The nonreserve countries agreed to fix their currencies to the U.S. dollar or to gold. More accurately, countries agreed to establish a “par value” exchange rate to the dollar and to maintain the exchange to within a 1 percent band around that par value. However, this detail is not an essential part of the story that follows. However, there was no obligation on the part of the nonreserve countries to exchange their currencies for gold. Only the reserve country had that obligation.

Instead, the nonreserve-currency countries were obliged to maintain the fixed exchange rate to the U.S. dollar by intervening on the foreign exchange (Forex) market and buying or selling dollars as necessary. In other words, when there was excess demand on the Forex for the home currency in exchange for dollars, the nonreserve central bank would supply their currency and buy dollars, thus running a balance of payments surplus, to maintain the fixity of their exchange rate.

Alternatively, when there was excess supply of the home currency, in exchange for dollars, the nonreserve central bank would supply dollars and buy its own currency on the Forex, resulting in a balance of payments deficit. Thus for all nonreserve countries the Bretton Woods system functioned like a reserve currency standard.

One of the problems that typically arises with a reserve currency standard is the persistence of balance of payments (BoP) deficits. BoP deficits require a country to sell its dollar reserves on the Forex market. When these deficits are recurring and large, a country will eventually run out of reserves. When that happens, it will no longer be able to defend its fixed currency value. The likely outcome would be a devaluation, an action that runs counter to the goals of the system, namely to maintain exchange rate stability and to ward off inflationary tendencies.

To provide a safety valve for countries that may face this predicament, the International Monetary Fund (IMF) was established to provide temporary loans to countries to help maintain their fixed exchange rates. Each member country was required to maintain a quota of reserves with the IMF that would then be available to lend to those countries experiencing balance of payments difficulties.

Today the IMF maintains the same quota system and member countries enjoy the same privilege to borrow even though many are no longer maintaining a fixed exchange rate. Instead, many countries borrow from the IMF when they become unable to maintain payments on international debts. Go to the [IMF Factsheet](#) for more information about the current quota system. International Monetary Fund, Factsheet, “IMF Quotas,” <http://www.imf.org/external/np/exr/facts/quotas.htm>

The Bretton Woods exchange rate system was an imperfect system that suffered under many strains during its history. Nonetheless, it did achieve fixed exchange rates among its members for almost thirty years. For a more detailed, though brief, account of the history of the system, see [Benjamin Cohen’s article](#). Benjamin Cohen, “Bretton Woods System,” <http://www.polsci.ucsb.edu/faculty/cohen/recent/bretton.html>.

We can learn much about the intended workings of the system by studying the system’s collapse. The collapse occurred mostly because the United States would not allow its internal domestic policies to be compromised for the sake of the fixed

exchange rate system. Here's a brief account of what happened. For a more detailed account, see Barry Eichengreen's *Globalizing Capital* (Princeton, NJ: Princeton University Press, 1996), and Alfred Eckes's *A Search for Solvency* (Austin, TX: University of Texas Press, 1975).

Throughout the 1960s and early 1970s, there was excessive supply of U.S. dollars on Forex markets in exchange for other currencies. This put pressure on the U.S. dollar to depreciate and nonreserve currencies to appreciate. To maintain the fixed exchange rate, nonreserve countries were required to intervene on the private Forex. For example, the British central bank was required to run a balance of payments surplus, buy the excess dollars, and sell pounds on the private Forex market.

As was shown in [Chapter 23 "Policy Effects with Fixed Exchange Rates", Section 23.6 "Currency Crises and Capital Flight"](#), persistent balance of payments surpluses do not pose a long-term problem in the same way as BoP deficits. The British central bank had an unlimited capacity to "print" as many pounds as necessary to buy the oversupplied dollars on the Forex. However, persistently large BoP surpluses will result in an ever-increasing British money supply that will lead to inflationary effects eventually.

Indeed, U.S. inflation was rising, especially in the late 1960s. Federal government spending was rising quickly—first, to finance the Vietnam War, and second, to finance new social spending arising out of President Johnson's Great Society initiatives. Rather than increasing taxes to finance the added expenses, the United States resorted to expansionary monetary policy, effectively printing money to finance growing government budget deficits. This is also called "monetizing the debt."

The immediate financial impact of a rising U.S. money supply was lower U.S. interest rates, leading to extra demand for foreign currency by investors to take advantage of the higher relative rates of return outside the United States. The longer-term impact of a rising U.S. money supply was inflation. As U.S. prices rose, U.S. goods became relatively more expensive relative to foreign goods, also leading to extra demand for foreign currency.

A look at the statistics of the 1960s belies this story of excessive monetary expansion and fiscal imprudence. Between 1959 and 1970, U.S. money supply growth and U.S. inflation were lower than in every other G-7 country. U.S. government budget deficits were also not excessively large. Nonetheless, as Eichengreen suggests, the G-7 countries could support a much higher inflation rate

than the United States since they were starting from such low levels of GDP in the wake of post-World War II reconstruction. Barry Eichengreen, *Globalizing Capital: A History of the International Monetary System* (Princeton, NJ: Princeton University Press, 1996), 131. Thus the U.S. policy required to maintain a stable exchange rate without intervention would correspond to an inflation rate that was considerably lower vis-à-vis the other G-7 countries.

In any case, to maintain the fixed exchange rate, non-U.S. countries' central banks needed to run balance of payments surpluses. BoP surpluses involved a nonreserve central bank purchase of dollars and sale of their own domestic currency. Thus the German, British, French, Japanese, et al., central banks bought up dollars in great quantities and at the same time continually increased their own domestic money supplies.

One effect of the continual balance of payments surpluses was a subsequent increase in inflation caused by rising money supplies in the nonreserve countries. In effect, expansionary monetary policy in the United States, and its inflationary consequences, are exported to the nonreserve countries by virtue of the fixed exchange rate system. This effect was not welcomed by the nonreserve countries like Britain, France, and Germany.

A second effect of the continual balance of payments surpluses was a rising stock of dollar reserves. Nonreserve central banks held those reserves in the form of U.S. Treasury bills; thus, increasingly, U.S. government debt was held by foreign countries.

Although such BoP surpluses could technically continue indefinitely, the inflationary consequences in Europe and Japan and the rising dollar holdings abroad put the sustainability of the system into question. Ideally in a fixed exchange system, BoP surpluses will be offset with comparable BoP deficits over time, if the exchange rate is fixed at an appropriate (i.e., sustainable) level. Continual BoP surpluses, however, indicate that the sustainable exchange rate should be at a much lower U.S. dollar value if the surpluses are to be eliminated. Recognition of this leads observers to begin to expect a dollar devaluation.

If (or when) a dollar devaluation occurred, dollar asset holdings by foreigners—including the U.S. government Treasury bills comprising the reserves held by foreign central banks—would suddenly fall in value. In other words, foreign asset holders would lose a substantial amount of money if the dollar were devalued.

For private dollar investors there was an obvious response to this potential scenario: divest of dollar assets—that is, sell dollars and convert to pounds, deutschmarks, or francs. This response in the late 1960s and early 1970s contributed to the capital flight from the U.S. dollar, put added downward pressure on the U.S. dollar value, and led to even greater BoP surpluses by nonreserve central banks.

The nonreserve central banks, on the other hand, could not simply convert dollars to pounds or francs, as this would add to the pressure for a depreciating dollar. Further, it was their dollar purchases that were preventing the dollar depreciation from happening in the first place.

During the 1960 and early 1970s the amount of U.S. dollar reserves held by nonreserve central banks grew significantly, which led to what became known as the **Triffin dilemma**⁶ (dollar overhang). Robert Triffin was a Belgian economist and Yale University professor who highlighted the problems related to dollar overhang. Dollar overhang occurred when the amount of U.S. dollar assets held by nonreserve central banks exceeded the total supply of gold in the U.S. Treasury at the exchange rate of \$35 per ounce. Dollar overhang occurred in the system by 1960 and continued to worsen throughout the decade of the 1960s. By 1971 foreign holdings of U.S. dollars stood at \$50 billion while U.S. gold reserves were valued at only \$15 billion.*Déclaration de Valéry Giscard d'Estaing à l'Assemblée nationale (12 mai 1971), dans La politique étrangère de la France. 1er semestre, octobre 1971, pp. 162–67.* Translated by le CVCE [Declaration by Valerie Giscard d'Estaing to the National Assembly (May 12, 1971)].

Under the Bretton Woods system, foreign central banks were allowed to exchange their dollars for gold at the rate of \$35 per ounce. Once the dollar overhang problem arose, it became conceivable that the United States could run out of its reserve asset—gold. Thus the potential for this type of BoP deficit could lead to speculation that the U.S. dollar would have to be devalued at some point in the future.

Now, if one expects the dollar will fall in value at some future date, then it would make sense to convert those dollars to something that may hold its value better; gold was the alternative asset. Throughout the 1950s and 1960s, foreign central banks did convert some of their dollar holdings to gold, but not all. In 1948, the United States held over 67 percent of the world's monetary gold reserves. By 1970, however, the U.S. gold holdings had fallen to just 16 percent of the world total.*Alfred E. Eckes Jr., A Search for Solvency* (Austin, TX: University of Texas Press, 1975), 238. In a gold exchange standard, the linkage between gold and the reserve currency is supposed to provide the constraint that prevents the reserve currency country from excessive monetary expansion and its subsequent inflationary effects.

6. The problem of excessive U.S. dollar holdings by foreign central banks.

However, in the face of BoP deficits leading to a severe depletion of gold reserves, the United States had several adjustment options open.

One option was a devaluation of the dollar. However, this option was not easy to implement. The U.S. dollar could not be devalued with respect to the pound, the franc, or the yen since the United States did not fix its currency to them. (Recall that the other countries were fixed to the dollar.) Thus the only way to realize this type of dollar devaluation was for the other countries to “revalue” their currencies with respect to the dollar. The other “devaluation” option open to the United States was devaluation with respect to gold. In other words, the United States could raise the price of gold to \$40 or \$50 per ounce or more. However, this change would not change the fundamental conditions that led to the excess supply of dollars. At most, this devaluation would only reduce the rate at which gold flowed out to foreign central banks. Also, since U.S. gold holdings had fallen to very low levels by the early 1970s and since the dollar overhang was substantial, the devaluation would have had to be extremely large to prevent the depletion of U.S. gold reserves.

The other option open to the United States was a change in domestic monetary policy to reduce the excess supply of dollars on the Forex. Recall that money supply increases were high to help finance rising federal deficit spending. A reversal of this policy would mean a substantial reduction in the growth of the money supply. If money supply increases were not available to finance the budget deficit, the government would have to resort to a much more unpopular method of financing—that is, raising taxes or reducing spending.

The unpopularity and internal difficulty of such fiscal and monetary prudence led the United States to resort to other options. One suggestion made repeatedly by the United States was that the nonreserve countries should “revalue” their currencies to the dollar. However, their response was that the fundamental problem was not their fault; therefore, they shouldn’t be the ones to implement a solution. Instead, it was the United States that needed to change.

By the spring of 1971, the imbalances in the system reached crisis proportions. In April 1971, the Bundesbank (Germany’s central bank) purchased over \$3 billion to maintain the fixed exchange rate. In early May, it bought over \$2 billion in just two days to maintain the rate. Fearing inflation after such huge purchases, Germany decided to let its currency float to a new value, 8 percent higher than its previous fixed rate. Austria, Holland, and Switzerland quickly followed suit. Alfred E. Eckes Jr., *A Search for Solvency* (Austin, TX: University of Texas Press, 1975), 261. Despite these revaluations, they were insufficient to stem the excess supply of dollars on the Forex. By August 1971, another major realignment seemed inevitable that substantially increased the pace of dollar capital flight. On August 15, 1971,

President Nixon announced a bold plan for readjustment. The plan had three main aspects:

1. A 10 percent import surcharge on all imports was implemented. This tariff would remain in effect until a new international monetary order was negotiated.
2. Suspension of dollar convertibility into gold. Foreign central banks would no longer have the option to exchange dollars for gold with the U.S. central bank.
3. Wage and price controls were implemented to stem the rising U.S. inflation

The import surcharge meant that an extra 10 percent would be assessed over the existing import tariff. This was implemented to force other countries to the bargaining table where, presumably, they would agree to a multilateral revaluation of their currencies to the dollar. The tax was especially targeted to pressure Japan, which had not revalued its currency as others had done during the previous years, to agree to a revaluation. The 10 percent import tax effectively raised the prices of foreign goods in U.S. markets and would have a similar effect as a 10 percent currency revaluation. The expectation was that the average revaluation necessary to bring the system into balance would be somewhat less than 10 percent, thus an 8 percent revaluation, say, would be less painful to exporters than a 10 percent import tax.

The suspension of dollar-gold convertibility was really the more significant change as it effectively ended the gold exchange standard and marked the death of the Bretton Woods system. With no obligation to exchange gold for dollars, the system essentially was changed to a reserve currency system. Previous constraints on the United States, caused when it runs a BoP deficit and loses gold reserves, were thus eliminated. There was no longer a possibility that the United States could run out of gold.

The wage and price controls, implemented for a ninety-day period, put added pressure on foreign exporters. Being forced to pay a 10 percent surcharge but not being allowed to raise prices meant they would not be allowed to push the tax increase onto consumers.

These three measures together resulted in a rapid renegotiation of the Bretton Woods system, culminating in the **Smithsonian Agreement**⁷ in December 1971. In this agreement, the nonreserve countries accepted an average 8 percent revaluation of their currencies to the dollar in return for the elimination of the import surcharge. They also enlarged the currency bands around the par values

7. A 1971 agreement meant to salvage the Bretton Woods system of fixed exchange rates.

from 1 percent to 2.25 percent. By virtue of the revaluations, the dollar naturally became “devalued.” The United States also devalued dollars with respect to gold, raising the price to \$38 per ounce. However, since the United States did not agree to reopen the gold window, the change in the price of gold was meaningless.

More important, since the United States no longer needed to be concerned about a complete loss of gold reserves, the dollar overhang problem was “solved,” and it was free to continue its monetary growth and inflationary policies. During the following year, the United States did just that; within a short time, there arose renewed pressure for the dollar to depreciate from its new par values.

In the end, the Smithsonian Agreement extended the life of Bretton Woods for just over a year. By March 1973, a repeat of the severe dollar outflows in 1971 led to a suspension of Forex trading for almost three weeks. Upon reopening, the major currencies were floating with respect to each other. The Bretton Woods system was dead.

The hope at the time was that floating rates could be allowed for a time to let exchange rates move to their market equilibrium rates. Once stability to the exchange rates was restored, a new fixed exchange rate system could be implemented. However, despite negotiations, an agreement was never reached, and a unified international system of fixed exchange rates has never since been tried.

How Bretton Woods Was Supposed to Work

In theory, a gold-exchange standard can work to provide exchange rate stability and reduce inflationary tendencies. However, it will only work if the reserve currency country maintains prudent monetary policies and if countries follow the rules of the system.

For the nonreserve countries, their task was to avoid balance of payments deficits. These deficits would arise if they pursued excessive expansionary monetary policy. The lower interest rates and eventual inflation would lead to capital flight, creating pressure for the currency to depreciate. To avoid a devaluation, and hence to follow the fixity rule, the nonreserve country would have to contract its money supply to take pressure off its currency and to reverse the BoP deficits.

The problem that usually arises here is that contractionary monetary policies will raise interest rates and eliminate an important source of government budget financing, namely debt monetization (printing money). These changes are likely to result in an increase in taxes, a decrease in government spending, a contraction of

the economy, and a loss of jobs. Thus following the rules of the system will sometimes be painful.

However, this was not the source of the Bretton Woods collapse. Instead, it was excessive monetary expansion by the reserve country, the United States. In this case, when the United States expanded its money supply, to finance budget deficits, it caused lower U.S. interest rates and had inflationary consequences. This led to increased demand for foreign currency by investors and traders. However, the United States was not obligated to intervene to maintain the fixed exchange rates since the United States was not fixing to anyone. Rather, it was the obligation of the nonreserve countries to intervene, buy dollars, sell their own currencies, and consequently run BoP surpluses. These surpluses resulted in the growing stock of dollar reserves abroad.

However, if the system had worked properly, foreign central banks would have cashed in their dollar assets for gold reserves long before the dollar overhang problem arose. With diminishing gold reserves, the United States would have been forced (i.e., if it followed the rules of the system) to reverse its expansionary monetary practices. However, as mentioned above, contractionary monetary policies will likely result in higher taxes, lower government spending, a contraction of the economy, and a loss of jobs.

Most countries faced with a choice between a policy that violates international monetary system rules and policies that maintain domestic vitality, even if only temporarily, will usually choose in favor of domestic interests. Of course, this choice will likely have negative longer-term consequences. Price and exchange rate stability will be compromised through these actions, and it will eliminate the benefits that would have come from expanded trade and international investments.

The gold exchange standard might have worked effectively if the United States and the others had committed themselves more intently on following the rules of the system. In the final analysis, what matters is the importance placed on maintaining the integrity of the cooperative fixed exchange rate system relative to the importance placed on domestic economic and political concerns. In the Bretton Woods case, domestic interests clearly dominated international interests.

The Bretton Woods experience should cast a shadow of doubt on fixed exchange rate systems more generally too. Every fixed exchange rate system requires countries to give up the independence of their monetary policy regardless of domestic economic circumstances. That this is difficult, or impossible, to do is demonstrated by the collapse of the Bretton Woods system.

KEY TAKEAWAYS

- The Bretton Woods system of exchange rates was set up as a gold exchange standard. The U.S. dollar was the reserve currency, and the dollar was fixed to gold at \$35 per ounce.
- The International Monetary Fund (IMF) was established to provide temporary loans to countries to help maintain their fixed exchange rates.
- U.S. expansionary monetary policy and its inflationary consequences were exported to the nonreserve countries by virtue of the fixed exchange rate system.
- The suspension of dollar-gold convertibility in 1971 effectively ended the gold exchange standard and marked the death of the Bretton Woods system.
- The Bretton Woods system collapsed in 1973 when all the currencies were allowed to float.
- A fixed exchange rate system requires nonreserve countries to give up the independence of their monetary policy regardless of domestic economic circumstances.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The Bretton Woods exchange system was this type of exchange rate standard.
 - b. The price of gold in terms of dollars when the Bretton Woods system began.
 - c. This international organization was created to help countries with balance of payments problems in the Bretton Woods system.
 - d. The percentage of world monetary gold held by the United States in 1948.
 - e. The percentage of world monetary gold held by the United States in 1970.
 - f. The name given to the problem of excessive U.S. dollar holdings by foreign central banks.
 - g. This country’s suspension of dollar convertibility to gold eliminated an important constraint that allowed the system to function properly.
 - h. The name of the agreement meant to salvage the Bretton Woods system in the early 1970s.
 - i. The month and year in which the Bretton Woods system finally collapsed.

Chapter 24

Fixed versus Floating Exchange Rates

One of the big issues in international finance is the appropriate choice of a monetary system. Countries can choose between a floating exchange rate system and a variety of fixed exchange rate systems. Which system is better is explored in this chapter. However, rather than suggesting a definitive answer, the chapter highlights the pros and cons of each type of system, arguing in the end that both systems can and have worked in some circumstances and failed in others.

24.1 Overview of Fixed versus Floating Exchange Rates

LEARNING OBJECTIVE

1. Preview the discussion about fixed versus floating exchange rate systems.

This chapter addresses what is perhaps the most important policy issue in international finance: to have fixed or floating exchange rates. The chapter focuses on three main features that affect the choice of system: volatility and risk, inflationary consequences, and monetary autonomy.

Volatility and risk refers to the tendency for exchange rates to change and the effect these changes have on the risk faced by traders and investors. Although in floating exchange systems volatility is a natural day-to-day occurrence, even in fixed exchange systems, devaluations or revaluations make volatility an issue. This chapter compares the two systems in light of this issue.

Inflationary consequences are shown to be a major potential problem for countries with floating exchange rates. For many countries facing this problem, fixed exchange rate systems can provide relief. The section shows that the relationship between inflation and the exchange rate system is an important element in the choice of system.

Finally, monetary autonomy, and the ability to control the economy, is lost with the choice of fixed exchange rates. We discuss why this loss of autonomy can be problematic in some circumstances but not in others.

The chapter concludes by providing some answers to the policy question, “fixed or floating?”

KEY TAKEAWAYS

- Three main features affect the choice of the exchange rate system: volatility and risk, inflationary consequences, and monetary autonomy.
- The choice between fixed and floating exchange rates is one of the most important policy decisions in international finance.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The choice between these two types of exchange rate systems is an important policy debate in international finance.
 - b. This term describing the extent to which an exchange rate may vary over time is an important consideration in the choice of exchange rate systems.
 - c. This term describing the likelihood of losing money is an important consideration in the choice of exchange rate systems.
 - d. Fixed exchange rates are sometimes chosen to mitigate this kind of general price problem.
 - e. This term describing the ability to influence the economy through monetary policy is an important consideration in the choice of exchange rate systems.

24.2 Exchange Rate Volatility and Risk

LEARNING OBJECTIVE

1. Learn how exchange rate volatility raises risk for international traders and investors.

Probably the most important characteristic of alternative exchange rate systems is the feature used to describe them, namely fixed or floating. Fixed exchange rates, by definition, are not supposed to change. They are meant to remain fixed, preferably permanently. Floating rates float up and down and down and up from year to year, week to week, and minute by minute. What a floating exchange rate will be a year from now, or even a week from now, is often very difficult to predict.

Volatility represents the degree to which a variable changes over time. The larger the magnitude of a variable change, or the more quickly it changes over time, the more volatile it is.

Since fixed exchange rates are not supposed to change—by definition—they have no volatility. Please note the cautious wording because fixed exchange rates are quite frequently devalued or revalued, implying that they can and do indeed change. However, we will explore this issue in more detail later. A floating exchange rate may or may not be volatile depending on how much it changes over time. However, since floating exchange rates are free to change, they are usually expected to be more volatile.

Volatile exchange rates make international trade and investment decisions more difficult because volatility increases exchange rate risk. **Exchange rate risk**¹ refers to the potential to lose money because of a change in the exchange rate. Below are two quick examples of how traders and investors may lose money when the exchange rate changes.

Exchange Rate Risk for Traders

1. Occurs when the profit in trade or the rate of return on an international investment can fall quickly because of a change in the exchange rate.

First consider a business that imports soccer balls into the United States. Suppose one thousand soccer balls purchased from a supplier in Pakistan costs 300,000 Pakistani rupees. At the current exchange rate of 60 Rs/\$, it will cost the importer \$5,000 dollars or \$5 per soccer ball. The importer determines that transportation,

insurance, advertising, and retail costs will run about \$5 per soccer ball. If the competitive market price for this type of soccer ball is \$12, he will make a \$2 profit per ball if all balls are sold.

Suppose the shipment is scheduled to occur in three months and that payment for the shipment need not be made until that time. Let's assume the importer waits to convert currency until the payment is made and that in three months' time the Pakistani rupee has appreciated to a new value of 55 Rs/\$. The shipment cost in rupees remains the same at Rs 300,000, but the dollar value of the shipment rises to \$5,454 or \$5.45 per soccer ball. Assuming the same \$5 of extra costs and a \$12 final sale price, the importer will now make only \$1.45 profit per soccer ball, if all balls are sold. While this is still a profit, it is about 25 percent less than expected when the decision to purchase was made three months before.

This is an example of the risk an importer faces because of a change in the currency value. Of course, it is true that the currency value could have changed in the opposite direction. Had the rupee value risen to 65 Rs/\$, the shipment value would have cost just \$4,615, or \$4.62 per ball, generating a profit of \$2.38 per soccer ball. In this case, the currency moves in the importer's favor. Thus a volatile exchange rate will sometimes lead to greater losses than expected, and at other times, to greater gains.

There are several methods to protect oneself from this type of currency risk. The importer could have exchanged currency at the time the deal was struck and held his 300,000 rupees in a Pakistani bank until payment is made. However, this involves a substantial additional opportunity cost since the funds must be available beforehand and become unusable while they are held in a Pakistani bank account. Alternatively, the importer may be able to find a bank willing to write a forward exchange contract, fixing an exchange rate today for an exchange to be made three months from now.

In any case, it should be clear that exchange rate fluctuations either increase the risk of losses relative to plans or increase the costs to protect against those risks.

Exchange Rate Risk for Investors

Volatile exchange rates also create exchange rate risk for international investors. Consider the following example. Suppose in October 2004, a U.S. resident decides to invest (i.e., save) \$10,000 for the next year. Given that the U.S. dollar had been weakening with respect to the Danish krone for several years and since the interest rate on a money market deposit was slightly higher in Denmark at 2.25 percent compared to the 1.90 percent return in the United States, the investor decides to

put the \$10,000 into the Danish account. At the time of the deposit, the exchange rate sits at 5.90 kr/\$. In October 2005, the depositor cashes in and converts the money back to U.S. dollars. The exchange rate in October 2005 was 6.23 kr/\$. To determine the return on the investment we can apply the rate of return formula derived in [Chapter 15 "Foreign Exchange Markets and Rates of Return"](#), [Section 15.3 "Calculating Rate of Returns on International Investments"](#) and [Chapter 15 "Foreign Exchange Markets and Rates of Return"](#), [Section 15.4 "Interpretation of the Rate of Return Formula"](#):

$$RoR_{kr} = i_{kr} + (1 + i_{kr}) \frac{\frac{1}{E_{kr/\$}^{2005}} - \frac{1}{E_{kr/\$}^{2004}}}{\frac{1}{E_{kr/\$}^{2004}}}$$

$$= .0225 + (1 + .0225) \frac{\frac{1}{6.23} - \frac{1}{5.90}}{\frac{1}{5.90}}$$

$$= -0.0317 \times 100 = -3.17\%$$

The rate of return works out to be negative, which means that instead of making money on the foreign deposit, this investor actually loses \$317. Had he deposited the \$10,000 in a U.S. account, he would have had a guaranteed return of 1.90 percent, earning him \$190 instead.

By depositing in a foreign account, the depositor subjected himself to exchange rate risk. The dollar unexpectedly appreciated during the year, resulting in a loss. Had the dollar remain fixed in value during that same time, the foreign return would have been 2.25 percent, which is larger than that obtained in the United States.

Thus fluctuating exchange rates make it more difficult for investors to know the best place to invest. One cannot merely look at what the interest rate is across countries but must also speculate about the exchange rate change. Make the wrong guess about the exchange rate movement and one could lose a substantial amount of money.

There are some ways to hedge against exchange rate risk. For example, with short-term deposits, an investor can purchase a forward contract or enter a futures market. In these cases, the investor would arrange to sell Danish krone in the future when the deposit is expected to be converted back to dollars. Since the future exchange rate is predetermined on such a contract, the rate of return is guaranteed

as well. Thus the risk of floating exchange rates can be reduced. However, for long-term investment such as foreign direct investment, these types of arrangements are more difficult and costly to implement.

Volatility and the Choice of Exchange Rate System

On the face of it, floating exchange rates would appear to be riskier than fixed rates since they are free to change regularly. For this reason, countries may choose fixed exchange rates to reduce volatility and thus to encourage international trade and investment.

The problem with this perception is that it has not worked out this way in practice. A 2004 International Monetary Fund (IMF) study by Peter Clark, Natalia Tamirisa, and Shang-Jin Wei, “Exchange Rate Volatility and Trade Flows—Some New Evidence,” International Monetary Fund, May 2004[0], <http://www.imf.org/external/np/res/exrate/2004/eng/051904.pdf>. notes that on average, during the 1970s, 1980s, and 1990s, the volatility of fixed exchange rates was approximately the same as that of floating rates. There are two reasons this can occur. First, a currency fixed to another reserve currency will continue to float against other currencies. Thus when China pegged its currency to the U.S. dollar, it continued to float with the dollar vis-à-vis the euro. Second, it is common for fixed currencies to be devalued or revalued periodically, sometimes dramatically. When this happens, the effects of volatility are concentrated in a very short time frame and can have much larger economic impacts.

The second thing noted by this study is that volatility had only a small effect on bilateral international trade flows, suggesting that the choice of exchange rate system on trade flows may be insignificant. However, the study does not consider the effects of volatility on international investment decisions. Other studies do show a negative relationship between exchange rate volatility and foreign direct investment. But if these results were true and fixed exchange rates are just as volatile as floating rates, then there is no obvious exchange system “winner” in terms of the effects on volatility. Nevertheless, volatility of exchange rate systems remains something to worry about and consider in the choice of exchange rate systems.

KEY TAKEAWAYS

- Volatile exchange rates make international trade and investment decisions more difficult because volatility increases exchange rate risk.
- Volatile exchange rates can quickly and significantly change the expected rates of return on international investments.
- Volatile exchange rates can quickly and significantly change the profitability of importing and exporting.
- Despite the expectation that fixed exchange rates are less volatile, a 2004 IMF study notes that on average, during the 1970s, 1980s, and 1990s, the volatility of fixed exchange rates was approximately the same as that of floating rates.

EXERCISES

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. This term describes the unpredictable movement of an exchange rate.
 - b. Of *increase, decrease, or no change*, the effect on an importer’s profits if he waits to exchange currency and the foreign currency rises in value vis-à-vis the domestic currency in the meantime.
 - c. Of *increase, decrease, or no change*, the effect on an importer’s profits if he waits to exchange currency and the domestic currency falls in value vis-à-vis the foreign currency in the meantime.
 - d. Of *increase, decrease, or no change*, the effect on an investor’s rate of return on foreign assets if the foreign currency rises in value more than expected vis-à-vis the domestic currency after purchasing a foreign asset.
 - e. Of *increase, decrease, or no change*, the effect on an investor’s rate of return on foreign assets if the foreign currency falls in value less than expected vis-à-vis the domestic currency after purchasing a foreign asset.
2. Between 2007 and 2008, the U.S. dollar depreciated significantly against the euro. Answer the following questions. Do not use graphs to explain. A one- or two-sentence verbal explanation is sufficient.
 - a. Explain whether European businesses that compete against U.S. imports gain or lose because of the currency change.
 - b. Explain whether European businesses that export their products to the United States gain or lose because of the currency change.
 - c. Explain whether European investors who purchased U.S. assets one year ago gain or lose because of the currency change.

24.3 Inflationary Consequences of Exchange Rate Systems

LEARNING OBJECTIVE

1. Learn how a fixed exchange rate system can be used to reduce inflation.

One important reason to choose a system of fixed exchange rates is to try to dampen inflationary tendencies. Many countries have (over time) experienced the following kind of situation. The government faces pressure from constituents to increase spending and raise transfer payments, which it does. However, it does not finance these expenditure increases with higher taxes since this is very unpopular. This leads to a sizeable budget deficit that can grow over time. When the deficits grow sufficiently large, the government may become unable to borrow additional money without raising the interest rate on bonds to unacceptably high levels. An easy way out of this fiscal dilemma is to finance the public deficits with purchases of bonds by the country's central bank. In this instance, a country will be financing the budget deficit by monetizing the debt, also known as printing money. New money means an increase in the domestic money supply, which will have two effects.

The short-term effect will be to lower interest rates. With free capital mobility, a reduction in interest rates will make foreign deposits relatively more attractive to investors and there is likely to be an increase in supply of domestic currency on the foreign exchange market. If floating exchange rates are in place, the domestic currency will depreciate with respect to other currencies. The long-term effect of the money supply increase will be inflation, if the gross domestic product (GDP) growth does not rise fast enough to keep up with the increase in money. Thus we often see countries experiencing a rapidly depreciating currency together with a rapid inflation rate. A good example of this trend was seen in Turkey during the 1980s and 1990s.

One effective way to reduce or eliminate this inflationary tendency is to fix one's currency. A fixed exchange rate acts as a constraint that prevents the domestic money supply from rising too rapidly. Here's how it works.

Suppose a country fixes its currency to another country—a reserve country. Next, imagine that the same circumstances from the story above begin to occur. Rising budget deficits lead to central bank financing, which increases the money supply of the country. As the money supply rises, interest rates decrease and investors begin

to move savings abroad, and so there is an increase in supply of the domestic currency on the foreign exchange market. However, now the country must prevent the depreciation of the currency since it has a fixed exchange rate. This means that the increase in supply of domestic currency by private investors will be purchased by the central bank to balance supply and demand at the fixed exchange rate. The central bank will be running a balance of payments deficit in this case, which will result in a reduction in the domestic money supply.

This means that as the central bank prints money to finance the budget deficit, it will simultaneously need to run a balance of payments deficit, which will soak up domestic money. The net effect on the money supply should be such as to maintain the fixed exchange rate with the money supply rising proportionate to the rate of growth in the economy. If the latter is true, there will be little to no inflation occurring. Thus a fixed exchange rate system can eliminate inflationary tendencies.

Of course, for the fixed exchange rate to be effective in reducing inflation over a long period, it will be necessary that the country avoid devaluations. Devaluations occur because the central bank runs persistent balance of payments deficits and is about to run out of foreign exchange reserves. Once the devaluation occurs, the country will be able to support a much higher level of money supply that in turn will have a positive influence on the inflation rate. If devaluations occur frequently, then it is almost as if the country is on a floating exchange rate system in which case there is no effective constraint on the money supply and inflation can again get out of control.

To make the fixed exchange rate system more credible and to prevent regular devaluation, countries will sometime use a currency board arrangement. With a currency board, there is no central bank with discretion over policy. Instead, the country legislates an automatic exchange rate intervention mechanism that forces the fixed exchange rate to be maintained.

For even more credibility, countries such as Ecuador and El Salvador have dollarized their currencies. In these cases, the country simply uses the other country's currency as its legal tender and there is no longer any ability to print money or let one's money supply get out of control.

However, in other circumstances fixed exchange rates have resulted in more, rather than less, inflation. In the late 1960s and early 1970s, much of the developed world was under the Bretton Woods system of fixed exchange rates. The reserve currency was the U.S. dollar, meaning that all other countries fixed their currency value to the U.S. dollar. When rapid increases in the U.S. money supply led to a surge of inflation in the United States, the other nonreserve countries like Britain, Germany,

France, and Japan were forced to run balance of payments surpluses to maintain their fixed exchange rates. These BoP surpluses raised these countries' money supplies, which in turn led to an increase in inflation. Thus, in essence, U.S. inflation was exported to many other countries because of the fixed exchange rate system.

The lesson from these stories is that sometimes fixed exchange rates tend to lower inflation while at other times they tend to increase it. The key is to fix your currency to something that is not likely to rise in value (inflate) too quickly. In the 1980 and 1990s, when the European **Exchange Rate Mechanism**² (ERM) was in place, countries were in practice fixed to the German deutschmark. Since the German central bank was probably the least prone to inflationary tendencies, all other European countries were able to bring their inflation rates down substantially due to the ERM system. However, had the countries fixed to the Italian lira, inflation may have been much more rapid throughout Europe over the two decades.

Many people propose a return to the gold standard precisely because it fixes a currency to something that is presumed to be steadier in value over time. Under a gold standard, inflation would be tied to the increase in monetary gold stocks. Because gold is strictly limited in physical quantity, only a limited amount can be discovered and added to gold stocks each year. Thus inflation may be adequately constrained. But because of other problems with a return to gold as the monetary support, a return to this type of system seems unlikely.

KEY TAKEAWAYS

- A fixed exchange rate can act as a constraint to prevent the domestic money supply from rising too rapidly (i.e., if the reserve currency country has noninflationary monetary policies).
- Adoption of a foreign country's currency as your own is perhaps the most credible method of fixing the exchange rate.
- Sometimes, as in the Bretton Woods system, a fixed exchange rate system leads to more inflation. This occurs if the reserve currency country engages in excessively expansionary monetary policy.
- A gold standard is sometimes advocated precisely because it fixes a currency to something (i.e., gold) that is presumed to be more steady in value over time.

2. An exchange rate system used within the European Union to maintain exchange rates within a specific band around a predetermined central exchange rate.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Hyperactivity in this aggregate variable is often a reason countries turn to fixed exchange rates.
 - b. If a country fixes its exchange rate, it effectively imports this policy from the reserve country.
 - c. A country fixing its exchange rate can experience high inflation if this country also experiences high inflation.
 - d. Of *relatively low* or *relatively high*, to limit inflation a country should choose to fix its currency to a country whose money supply growth is this.
 - e. The name for the post–World War II exchange rate system that demonstrated how countries fixing their currency could experience high inflation.

24.4 Monetary Autonomy and Exchange Rate Systems

LEARNING OBJECTIVE

1. Learn how floating and fixed exchange rate systems compare with respect to monetary autonomy.

Monetary autonomy³ refers to the independence of a country's central bank to affect its own money supply and conditions in its domestic economy. In a floating exchange rate system, a central bank is free to control the money supply. It can raise the money supply when it wishes to lower domestic interest rates to spur investment and economic growth. By doing so it may also be able to reduce a rising unemployment rate. Alternatively, it can lower the money supply, to raise interest rates and to try to choke off excessive growth and a rising inflation rate. With monetary autonomy, monetary policy is an available tool the government can use to control the performance of the domestic economy. This offers a second lever of control, beyond fiscal policy.

In a fixed exchange rate system, monetary policy becomes ineffective because the fixity of the exchange rate acts as a constraint. As shown in [Chapter 23 "Policy Effects with Fixed Exchange Rates"](#), [Section 23.2 "Monetary Policy with Fixed Exchange Rates"](#), when the money supply is raised, it will lower domestic interest rates and make foreign assets temporarily more attractive. This will lead domestic investors to raise demand for foreign currency that would result in a depreciation of the domestic currency, if a floating exchange rate were allowed. However, with a fixed exchange rate in place, the extra demand for foreign currency will need to be supplied by the central bank, which will run a balance of payments deficit and buy up its own domestic currency. The purchases of domestic currency in the second stage will perfectly offset the increase in money in the first stage, so that no increase in money supply will take place.

Thus the requirement to keep the exchange rate fixed constrains the central bank from using monetary policy to control the economy. In other words, the central bank loses its autonomy or independence.

3. Refers to the independence of a country's central bank to affect its own money supply and conditions in its domestic economy.

In substitution, however, the government does have a new policy lever available in a fixed system that is not available in a floating system, namely exchange rate policy. Using devaluations and revaluations, a country can effectively raise or lower the money supply level and affect domestic outcomes in much the same way as it

might with monetary policy. However, regular exchange rate changes in a fixed system can destroy the credibility in the government to maintain a truly “fixed” exchange rate. This in turn could damage the effect fixed exchange rates might have on trade and investment decisions and on the prospects for future inflation.

Nonetheless, some countries do apply a semifixed or semifloating exchange rate system. A crawling peg, in which exchange rates are adjusted regularly, is one example. Another is to fix the exchange rate within a band. In this case, the central bank will have the ability to control the money supply, up or down, within a small range, but will not be free to make large adjustments without breaching the band limits on the exchange rate. These types of systems provide an intermediate degree of autonomy for the central bank.

If we ask which is better, monetary autonomy or a lack of autonomy, the answer is mixed. In some situations, countries need, or prefer, to have monetary autonomy. In other cases, it is downright dangerous for a central bank to have autonomy. The determining factor is whether the central bank can maintain prudent monetary policies. If the central bank can control money supply growth such that it has only moderate inflationary tendencies, then monetary autonomy can work well for a country. However, if the central bank cannot control money supply growth, and if high inflation is a regular occurrence, then monetary autonomy is not a blessing.

One of the reasons Britain has decided not to join the eurozone is because it wants to maintain its monetary autonomy. By joining the eurozone, Britain would give up its central bank’s ability to control its domestic money supply since euros would circulate instead of British pounds. The amount of euros in circulation is determined by the European Central Bank (ECB). Although Britain would have some input into money supply determinations, it would clearly have much less influence than it would for its own currency. The decisions of the ECB would also reflect the more general concerns of the entire eurozone rather than simply what might be best for Britain. For example, if there are regional disparities in economic growth (e.g., Germany, France, etc., are growing rapidly, while Britain is growing much more slowly), the ECB may decide to maintain a slower money growth policy to satisfy the larger demands to slow growth and subsequent inflation in the continental countries. The best policy for Britain alone, however, might be a more rapid increase in money supply to help stimulate its growth. If Britain remains outside the eurozone, it remains free to determine the monetary policies it deems best for itself. If it joins the eurozone, it loses its monetary autonomy.

In contrast, Argentina suffered severe hyperinflations during the 1970s and 1980s. Argentina’s central bank at the time was not independent of the rest of the national government. To finance large government budget deficits, Argentina resorted to

running the monetary printing presses, which led to the severe hyperinflations. In this case, monetary autonomy was a curse, not a blessing.

In an attempt to restrain the growth of the money supply, Argentina imposed a currency board in 1992. A currency board is a method of fixing one's exchange rate with a higher degree of credibility. By legislating mandatory automatic currency interventions, a currency board operates in place of a central bank and effectively eliminates the autonomy that previously existed. Although Argentina's currency board experiment collapsed in 2002, for a decade Argentina experienced the low inflation that had been so elusive during previous decades.

KEY TAKEAWAYS

- Monetary autonomy refers to the independence of a country's central bank to affect its own money supply and, through that, conditions in its domestic economy.
- In a fixed exchange rate system, a country maintains the same interest rate as the reserve country. As a result, it loses the ability to use monetary policy to control outcomes in its domestic economy.
- In a floating exchange rate system, a country can adjust its money supply and interest rates freely and thus can use monetary policy to control outcomes in its domestic economy.
- If the central bank can control money supply growth such that it has only moderate inflationary tendencies, then monetary autonomy (floating) can work well for a country. However, if the central bank cannot control money supply growth, and if high inflation is a regular occurrence, then monetary autonomy (floating) will not help the country.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. The term describing the relationship between the U.S. Federal Reserve Board and the U.S. government that has quite likely contributed to the low U.S. inflation rate in the past two decades.
 - b. In part to achieve this, the United Kingdom has refused to adopt the euro as its currency.
 - c. Of *fixed* or *floating*, in this system a country can effectively set its money supply at any level desired.
 - d. Of *fixed* or *floating*, in this system a country’s interest rate will always be the same as the reserve country’s.
 - e. Of *fixed* or *floating*, in this system a country can control inflation by maintaining moderate money supply growth.

24.5 Which Is Better: Fixed or Floating Exchange Rates?

LEARNING OBJECTIVE

1. Learn the pros and cons of both floating and fixed exchange rate systems.

The exchange rate is one of the key international aggregate variables studied in an international finance course. It follows that the choice of exchange rate system is one of the key policy questions.

Countries have been experimenting with different international payment and exchange systems for a very long time. In early history, all trade was barter exchange, meaning goods were traded for other goods. Eventually, especially scarce or precious commodities, for example gold and silver, were used as a medium of exchange and a method for storing value. This practice evolved into the metal standards that prevailed in the nineteenth and early twentieth centuries. By default, since gold and silver standards imply fixed exchange rates between countries, early experience with international monetary systems was exclusively with fixed systems. Fifty years ago, international textbooks dealt almost entirely with international adjustments under a fixed exchange rate system since the world had had few experiences with floating rates.

That experience changed dramatically in 1973 with the collapse of the Bretton Woods fixed exchange rate system. At that time, most of the major developed economies allowed their currencies to float freely, with exchange values being determined in a private market based on supply and demand, rather than by government decree. Although when Bretton Woods collapsed, the participating countries intended to resurrect a new improved system of fixed exchange rates, this never materialized. Instead, countries embarked on a series of experiments with different types of fixed and floating systems.

For example, the European Economic Community (now the EU) implemented the exchange rate mechanism in 1979, which fixed each other's currencies within an agreed band. These currencies continued to float with non-EU countries. By 2000, some of these countries in the EU created a single currency, the euro, which replaced the national currencies and effectively fixed the currencies to each other immutably.

Some countries have fixed their currencies to a major trading partner, and others fix theirs to a basket of currencies comprising several major trading partners. Some have implemented a crawling peg, adjusting the exchange values regularly. Others have implemented a dirty float where the currency value is mostly determined by the market but periodically the central bank intervenes to push the currency value up or down depending on the circumstances. Lastly, some countries, like the United States, have allowed an almost pure float with central bank interventions only on rare occasions.

Unfortunately, the results of these many experiments are mixed. Sometimes floating exchange rate systems have operated flawlessly. At other times, floating rates have changed at breakneck speed, leaving traders, investors, and governments scrambling to adjust to the volatility. Similarly, fixed rates have at times been a salvation to a country, helping to reduce persistent inflation. At other times, countries with fixed exchange rates have been forced to import excessive inflation from the reserve country.

No one system has operated flawlessly in all circumstances. Hence, the best we can do is to highlight the pros and cons of each system and recommend that countries adopt that system that best suits its circumstances.

Probably the best reason to adopt a fixed exchange rate system is to commit to a loss in monetary autonomy. This is necessary whenever a central bank has been independently unable to maintain prudent monetary policy, leading to a reasonably low inflation rate. In other words, when inflation cannot be controlled, adopting a fixed exchange rate system will tie the hands of the central bank and help force a reduction in inflation. Of course, in order for this to work, the country must credibly commit to that fixed rate and avoid pressures that lead to devaluations. Several methods to increase the credibility include the use of currency boards and complete adoption of the other country's currency (i.e., dollarization or euroization). For many countries, for at least a period, fixed exchange rates have helped enormously to reduce inflationary pressures.

Nonetheless, even when countries commit with credible systems in place, pressures on the system sometimes can lead to collapse. Argentina, for example, dismantled its currency board after ten years of operation and reverted to floating rates. In Europe, economic pressures have led to some “talk” about giving up the euro and returning to national currencies. The Bretton Woods system lasted for almost thirty years but eventually collapsed. Thus it has been difficult to maintain a credible fixed exchange rate system for a long period.

Floating exchange rate systems have had a similar colored past. Usually, floating rates are adopted when a fixed system collapses. At the time of a collapse, no one really knows what the market equilibrium exchange rate should be, and it makes some sense to let market forces (i.e., supply and demand) determine the equilibrium rate. One of the key advantages of floating rates is the autonomy over monetary policy that it affords a country's central bank. When used wisely, monetary policy discretion can provide a useful mechanism for guiding a national economy. A central bank can inject money into the system when the economic growth slows or falls, or it can reduce money when excessively rapid growth leads to inflationary tendencies. Since monetary policy acts much more rapidly than fiscal policy, it is a much quicker policy lever to use to help control the economy.

Prudent Monetary and Fiscal Policies

Interestingly, monetary autonomy is both a negative trait for countries choosing fixed rates to rid themselves of inflation and a positive trait for countries wishing have more control over their domestic economies. It turns out that the key to success in both fixed and floating rates hinges on prudent monetary and fiscal policies. Fixed rates are chosen to force a more prudent monetary policy, while floating rates are a blessing for those countries that already have a prudent monetary policy.

A prudent monetary policy is most likely to arise when two conditions are satisfied. First, the central bank, and the decisions it makes, must be independent of the national government that makes government-spending decisions. If it is not, governments have always been inclined to print money to finance government-spending projects. This has been the primary source of high inflation in most countries. The second condition is a clear guideline for the central bank's objective. Ideally, that guideline should broadly convey a sense that monetary policy will satisfy the demands of a growing economy while maintaining sufficiently low inflation. When these conditions are satisfied, autonomy for a central bank and floating exchange rates will function well. Mandating fixed exchange rates can also work well, but only if the system can be maintained and if the country to which the other country fixes its currency has a prudent monetary policy.

Both systems can experience great difficulties if prudent fiscal policies are not maintained. This requires governments to maintain a balanced budget over time. Balance over time does not mean balance in every period but rather that periodic budget deficits should be offset with periodic budget surpluses. In this way, government debt is managed and does not become excessive. It is also critical that governments do not overextend themselves in terms of international borrowing. International debt problems have become the bane of many countries.

Unfortunately, most countries have been unable to accomplish this objective. Excessive government deficits and borrowing are the norm for both developing and developed countries. When excessive borrowing needs are coupled with a lack of central bank independence, tendencies to hyperinflations and exchange rate volatility are common. When excessive borrowing is coupled with an independent central bank and a floating exchange rate, exchange rate volatility is also common.

Stability of the international payments system then is less related to the type of exchange rate system chosen than it is to the internal policies of the individual countries. Prudent fiscal and monetary policies are the keys.

With prudent domestic policies in place, a floating exchange rate system will operate flawlessly. Fixed exchange systems are most appropriate when a country needs to force itself to a more prudent monetary policy course.

KEY TAKEAWAYS

- Historically, no one system has operated flawlessly in all circumstances.
- Probably the best reason to adopt a fixed exchange rate system is whenever a central bank has been independently unable to maintain prudent monetary policy, leading to a reasonably low inflation rate.
- Probably the best reason to adopt a floating exchange rate system is whenever a country has more faith in the ability of its own central bank to maintain prudent monetary policy than any other country's ability.
- The key to success in both fixed and floating rates hinges on prudent monetary and fiscal policies. Fixed rates are chosen to force a more prudent monetary policy; floating rates are a blessing for those countries that already have a prudent monetary policy.

EXERCISE

1. **Jeopardy Questions.** As in the popular television game show, you are given an answer to a question and you must respond with the question. For example, if the answer is “a tax on imports,” then the correct question is “What is a tariff?”
 - a. Of *fixed* or *floating*, this system is often chosen by countries that in their recent history experienced very high inflation.
 - b. Of *fixed* or *floating*, this system is typically chosen when a country has confidence in its own ability to conduct monetary policy effectively.
 - c. Of *fixed* or *floating*, this system is typically chosen when a country has little confidence in its own ability to conduct monetary policy effectively.
 - d. Of *fixed* or *floating*, this system is sometimes rejected because it involves the loss of national monetary autonomy.
 - e. Of *fixed* or *floating*, this system is sometimes chosen because it involves the loss of national monetary autonomy.