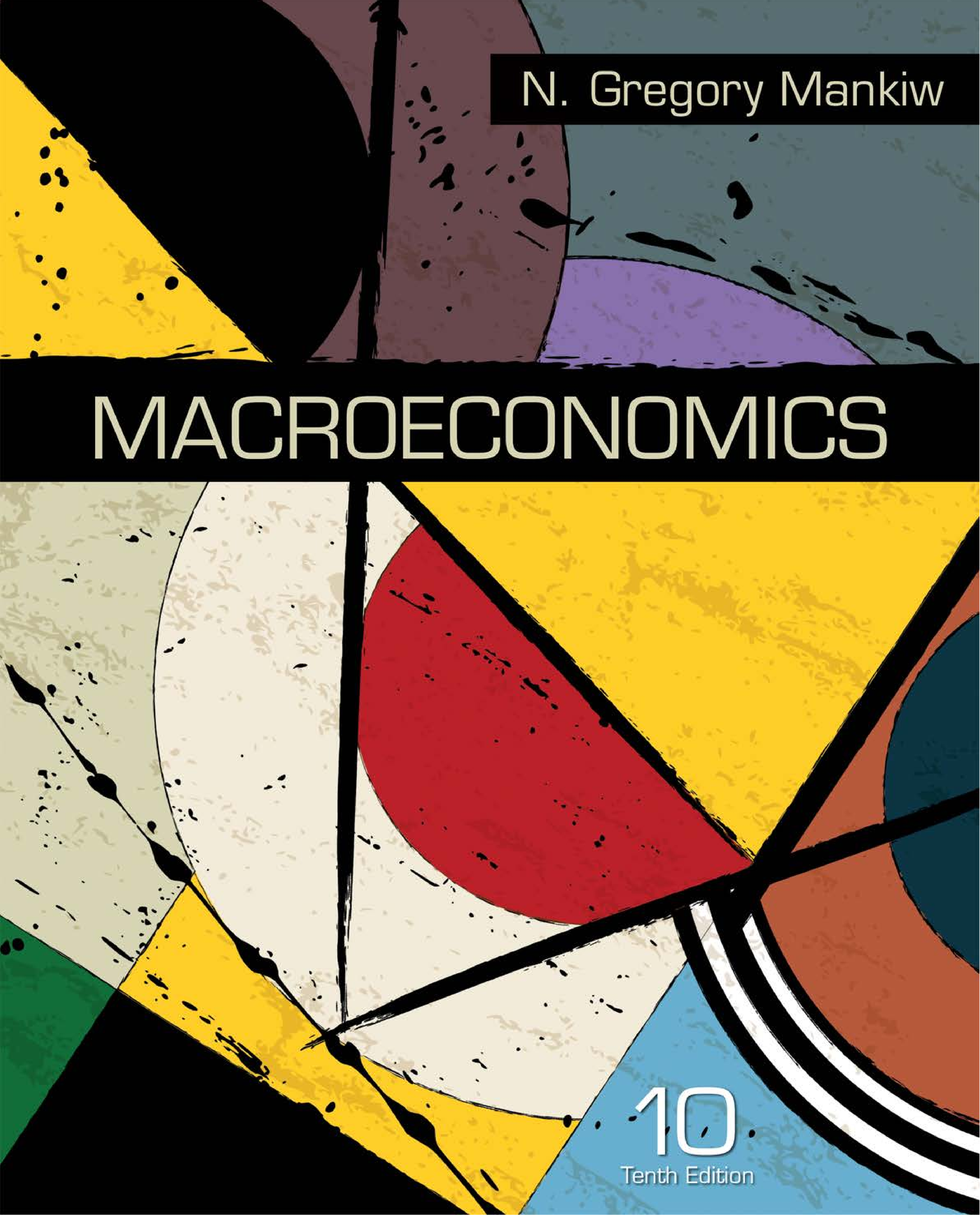


N. Gregory Mankiw

MACROECONOMICS

10

Tenth Edition

MACROECONOMICS

TENTH EDITION

MACROECONOMICS

N. GREGORY MANKIW
Harvard University

worth publishers
Macmillan Learning
New York

Senior Vice President, Content Strategy: Charles Linsmeier
Program Director, Social Sciences: Shani Fisher
Executive Program Manager: Simon Glick
Development Editor: Jane E. Tufts
Assistant Editor: Courtney Lindwall
Editorial Assistant: Amanda Gaglione
Marketing Manager: Andrew Zierman
Marketing Assistant: Chelsea Simens
Director of Media Editorial & Assessment, Social Sciences: Noel Hohnstine
Associate Media Editor: Nikolas Toner
Assessment Manager: Kristyn Brown
Assessment Editor: Joshua Hill
Media Project Manager: Andrew Vaccaro
Director, Content Management Enhancement: Tracey Kuehn
Senior Managing Editor: Lisa Kinne
Senior Photo Editor: Robin Fadool
Director of Design, Content Management: Diana Blume
Design Services Manager: Natasha A. S. Wolfe
Cover Designer: John Callahan
Text Designer: Kall Design
Senior Workflow Manager: Susan Wein
Composition: Lumina Datamatics, Inc.
Cover Art: Kirsten Hinte/Shutterstock

Library of Congress Control Number: 2018941309

ISBN: 978-1-319-10605-8 (epub)

1 2 3 4 5 6 23 22 21 20 19 18

Worth Publishers
One New York Plaza
Suite 4500
New York, NY 10004-1562
www.macmillanlearning.com

About the Author

Jordi Cabré

N. Gregory Mankiw is the Robert M. Beren Professor of Economics at Harvard University. He began his study of economics at Princeton University, where he received an A.B. in 1980. After earning a Ph.D. in economics from MIT, he began teaching at Harvard in 1985 and was promoted to full professor in 1987. At -Harvard, he has taught both undergraduate and graduate courses in macroeconomics. He is also author of the best-selling introductory textbook *Principles of Economics* (Cengage Learning).

Professor Mankiw is a regular participant in academic and policy debates. His research ranges across macroeconomics and includes work on price adjustment, consumer behavior, financial markets, monetary and fiscal policy, and economic growth. In addition to his duties at Harvard, he has been a research associate of the National Bureau of Economic Research, a member of the Brookings Panel on Economic Activity, a trustee of the Urban Institute, and an adviser to the Congressional Budget Office and the Federal Reserve Banks of Boston and New York. From 2003 to 2005, he was chair of the President’s Council of Economic Advisers.

Professor Mankiw lives in Massachusetts with his wife, Deborah, and their children, Catherine, Nicholas, and Peter.

To Deborah

Those branches of politics, or of the laws of social life, in which there exists a collection of facts or thoughts sufficiently sifted and methodized to form the beginning of a science should be taught *ex professo*. Among the chief of these is Political Economy, the sources and conditions of wealth and material prosperity for aggregate bodies of human beings. . . .

The same persons who cry down Logic will generally warn you against Political Economy. It is unfeeling, they will tell you. It recognises unpleasant facts. For my part, the most unfeeling thing I know of is the law of gravitation: it breaks the neck of the best and most amiable person without scruple, if he forgets for a single moment to give heed to it. The winds and waves too are very unfeeling. Would you advise those who go to sea to deny the winds and waves—or to make use of them, and find the means of guarding against their dangers? My advice to you is to study the great writers on Political Economy, and hold firmly by whatever in them you find true; and depend upon it that if you are not selfish or hardhearted already, Political Economy will not make you so.

John Stuart Mill, 1867

Brief Contents

Media and Resources from Worth Publishers

Digital Resources for Students and Instructors

SaplingPlus

Worth Publishers' new online course system, SaplingPlus, combines Learning-Curve with an integrated e-book, robust homework, improved graphing, and fully digital end-of-chapter problems, including Work It Outs. Online homework helps students get better grades with targeted instructional feedback tailored to the individual. And it saves instructors time preparing for and managing a course by providing personalized support from a Ph.D. or master's level colleague trained in Sapling's system.

Worth Publishers has worked closely with Greg Mankiw and a team of talented economics instructors to assemble a variety of resources for instructors and students. We have been delighted by all of the positive feedback we have received.

For Instructors

Instructor's Resource Manual

Robert G. Murphy (Boston College) has revised the impressive resource manual for instructors. For each chapter of this book, the manual contains notes to the instructor, a detailed lecture outline, additional case studies, and coverage of advanced topics. Instructors can use the manual to prepare their lectures, and they can reproduce whatever pages they choose as handouts for students. Each chapter also contains a Moody's Analytics Economy.com Activity (www.economy.com), which challenges students to combine the chapter knowledge with a high-powered business database and analysis service that offers real-time monitoring of the global economy.

Solutions Manual

Mark Gibson (Washington State University) has updated the *Solutions Manual* for all the Questions for Review and Problems and Applications found in the text.

Test Bank

The Test Bank has been extensively revised and improved for the tenth edition. Based on reviewer feedback, Worth Publishers, in collaboration with Daniel Moncayo (Brigham Young University), has checked every question, retained only the best, and added fresh new questions. The Test Bank now includes more than 2,200 multiple-choice questions, numerical problems, and short-answer graphical questions to accompany each chapter. The Test Bank provides a wide range of questions appropriate for assessing students' comprehension, interpretation, analysis, and synthesis skills.

Lecture Slides

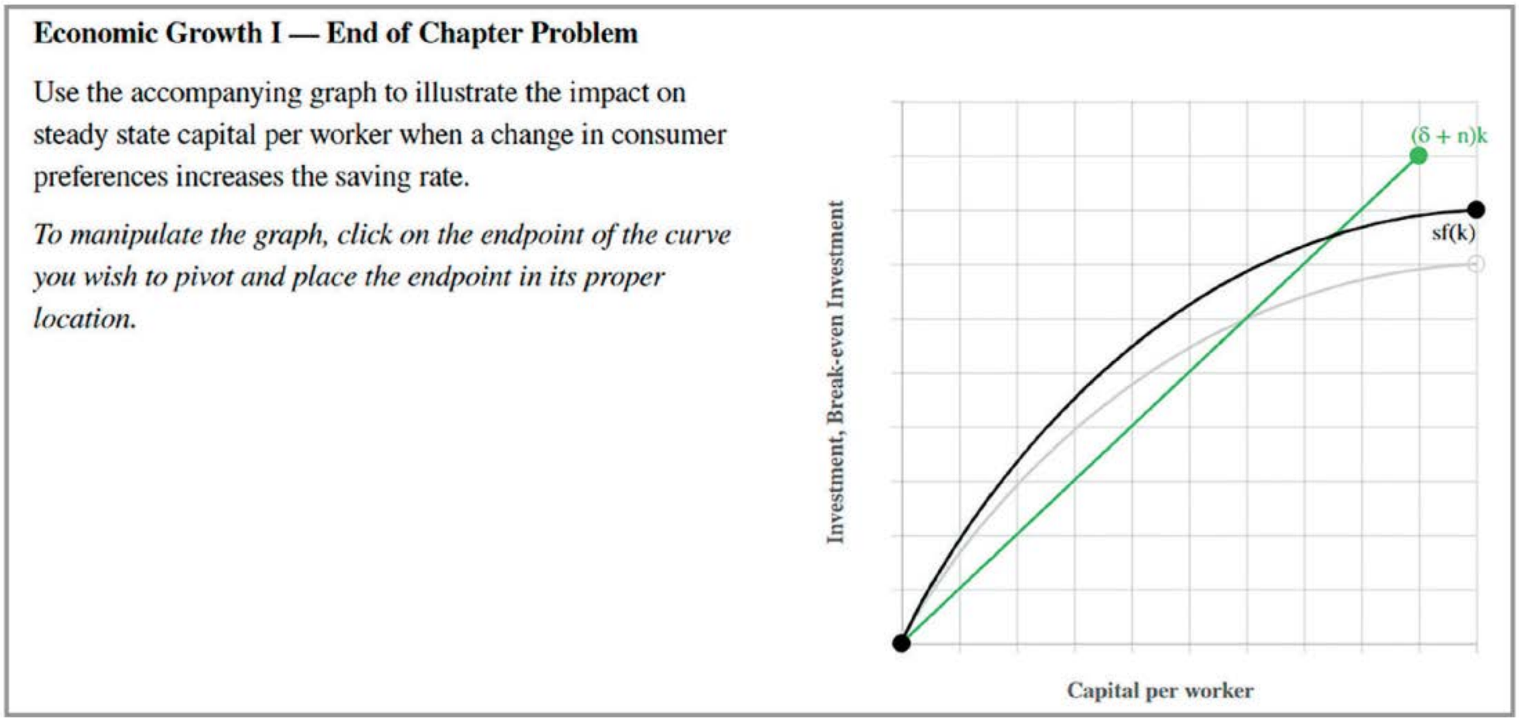
Ryan Lee (Indiana University) has revised his lecture slides for the material in each chapter. They feature graphs with careful explanations and additional case studies, data, and helpful notes to the instructor. Designed to be customized or used as is, the slides include easy directions for instructors who have little PowerPoint experience.

End-of-Chapter Problems

The end-of-chapter problems from the text have been converted to an interactive format with answer-specific feedback. These problems can be assigned as homework assignments or in quizzes.

Graphing Questions

Powered by improved graphing, multi-step questions paired with helpful feedback guide students through the process of problem solving. Students are asked to demonstrate their understanding by simply clicking, dragging, and dropping a line to a predetermined location. The graphs have been designed so that students' entire focus is on moving the correct curve in the correct direction, virtually eliminating grading issues for instructors.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

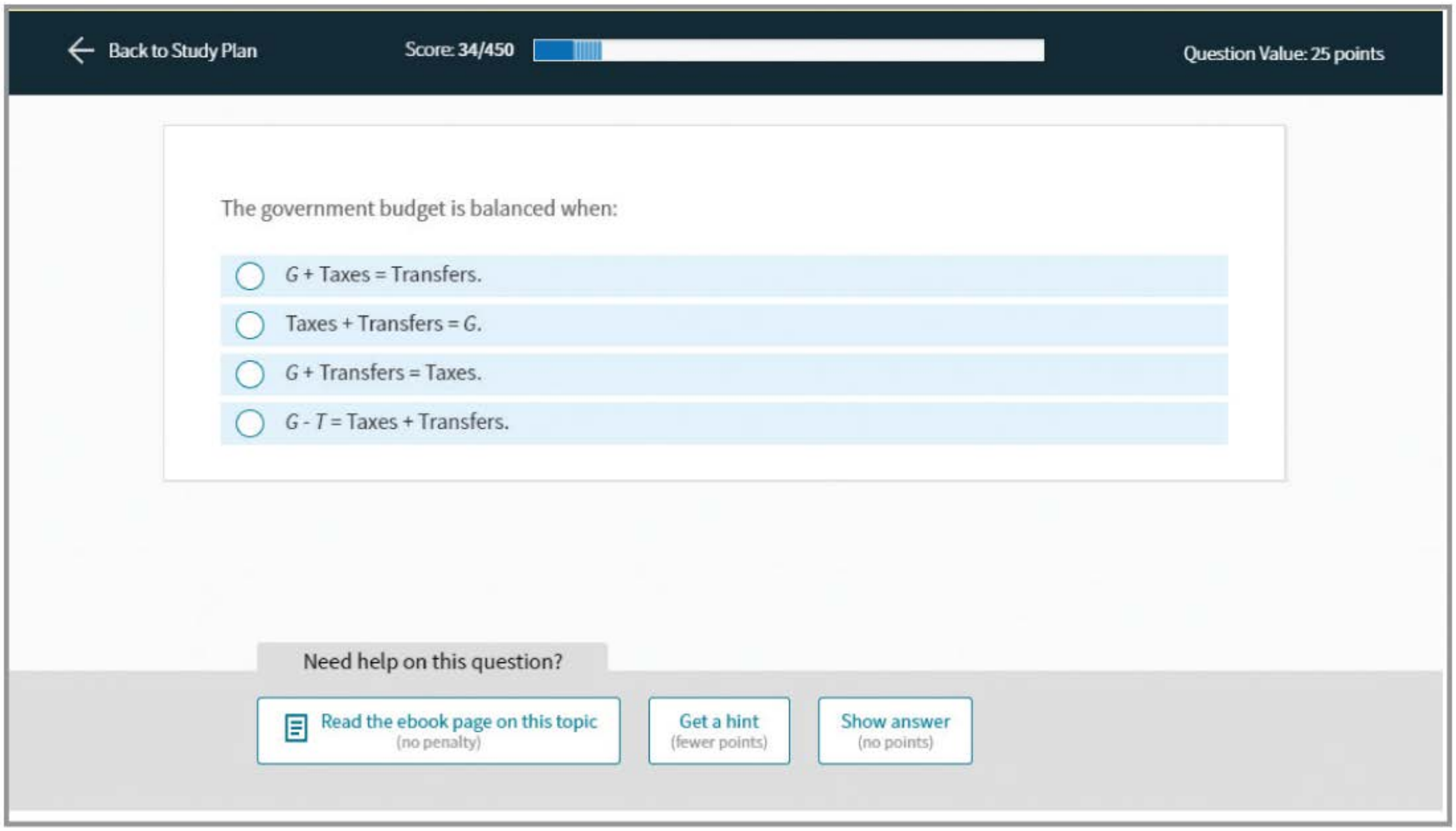
Homework Assignments

Each chapter contains prebuilt assignments, providing instructors with a curated set of multiple-choice and graphing questions that can be easily assigned for practice or graded assessments.

For Students

LearningCurve

LearningCurve is an adaptive quizzing engine that automatically adjusts questions to a student’s mastery level. With LearningCurve activities, each student follows a unique path to understanding the material. The more questions a student answers correctly, the more difficult the questions become. Each question is written specifically for the text and is linked to the relevant e-book section. LearningCurve also provides a personal study plan for students as well as complete metrics for instructors. LearningCurve, which has been proved to raise student performance, serves as an ideal formative assessment and learning tool.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

Work It Out Tutorials

These skill-building activities pair sample end-of-chapter problems (identified with this icon: **Work It Out**) with targeted feedback and video explanations to help students solve a similar problem step by step. This approach allows students to work independently, tests their comprehension of concepts, and prepares them for class and exams.

Fed Chairman Game

Created by the Federal Reserve Bank of San Francisco, the game allows students to become Chairman of the Fed and to make macroeconomic policy decisions based on news events and economic statistics. This fun-to-play simulation gives students a sense of the complex interconnections that influence the economy.

Prelude: Celebrating the Tenth Edition

I started writing the first edition of this book in 1988. My department chair had asked me to teach intermediate macroeconomics, a required course for Harvard economics majors. I happily accepted the assignment and continued teaching intermediate macro for the next 15 years. (I stepped away only when asked to take over the principles course.) As I prepared for the course by surveying existing texts, I realized that none of them fully satisfied me. While many were excellent books, I felt that they did not provide the right balance between long-run and short-run perspectives, between classical and Keynesian insights. And some were too long and comprehensive to be easily taught in one semester. Thus, this book was born.

Since its initial publication in 1991, the book has found an eager audience. My publisher tells me that it has been the best-selling intermediate macroeconomics text throughout most of its life. That is truly heartening. I am grateful to the numerous instructors who have adopted the book and, over many editions, have helped me improve it with their input. Even more heartening are the letters and emails from students around the world, who tell me how the book helped them navigate the exciting and challenging field of macroeconomics.

Over the past 30 years, macroeconomics has evolved as history has presented new questions and research has offered new answers. When the first edition came out, no one had heard of digital currencies such as bitcoin, Europe did not have a common currency, John Taylor had not devised his eponymous rule for monetary policy, behavioral economists like David Laibson and Richard Thaler had not proposed new ways to explain consumer behavior, and the economics profession had yet to be forced by the events of 2008 to focus anew on financial crises. Because of these and many other developments, I have updated this book every three years to ensure that students always have access to state-of-the-art thinking.

We macroeconomists still have much to learn. But the current body of macroeconomic knowledge offers students much insight into the world in which they live. Nothing delights me more than knowing that this book has helped convey this insight to the next generation.

Preface

An economist must be "mathematician, historian, statesman, philosopher, in some degree . . . as aloof and incorruptible as an artist, yet sometimes as near to earth as a politician." So remarked John Maynard Keynes, the great British economist who could be called the father of macroeconomics. No single statement summarizes better what it means to be an economist.

As Keynes suggests, students learning economics must draw on many disparate talents. The job of helping students develop these talents falls to instructors and textbook authors. My goal for this book is to make macroeconomics understandable, relevant, and (believe it or not) fun. Those of us who have chosen to be macroeconomists have done so because we are fascinated by the field. More important, we believe that the study of macroeconomics can illuminate much about the world and that the lessons learned, if properly applied, can make the world a better place. I hope this book conveys not only our profession's wisdom but also its enthusiasm and sense of purpose.

This Book's Approach

Macroeconomists share a common body of knowledge, but they do not all have the same perspective on how that knowledge is best taught. Let me begin this new edition by recapping my objectives, which together define this book's approach to the field.

First, I try to offer a balance between short-run and long-run topics. All economists agree that public policies and other events influence the economy over different time horizons. We live in our own short run, but we also live in the long run that our parents bequeathed us. As a result, courses in macroeconomics need to cover both short-run topics, such as the business cycle and stabilization policy, and long-run topics, such as economic growth, the natural rate of unemployment, persistent inflation, and the effects of government debt. Neither time horizon trumps the other.

Second, I integrate the insights of Keynesian and classical theories. Keynes's *General Theory* is the foundation for much of our understanding of economic fluctuations, but classical economics provides the right answers to many questions. In this book I incorporate the contributions of the classical economists before Keynes and the new classical economists of the past several decades. Substantial coverage is given, for example, to the loanable-funds theory of the interest rate, the quantity theory of money, and the problem of time inconsistency. At the same time, the ideas of Keynes and the new Keynesians are necessary for understanding fluctuations. Substantial coverage is also given to the *IS–LM* model of aggregate demand, the short-run tradeoff between inflation and unemployment, and modern models of business cycle dynamics.

Third, I present macroeconomics using a variety of simple models. Instead of pretending that there is one model complete enough to explain all facets of the economy, I encourage students to learn how to use a set of prominent models. This approach has the pedagogical value that each model can be kept simple and presented within one or two chapters. More important, this approach asks students to think like economists, who always keep various models in mind when analyzing economic events or public policies.

Fourth, I emphasize that macroeconomics is an empirical discipline, motivated and guided by a wide array of experience. This book contains numerous case studies that use macroeconomic theory to shed light on real-world data and events. To highlight the broad applicability of the theory, I have drawn the case studies both from current issues facing the world's economies and from dramatic historical episodes. They teach the reader how to apply economic principles to issues from fourteenth-century Europe, the island of Yap, the land of Oz, and today's newspaper.

What's New in the Tenth Edition?

Here is a brief rundown of the notable changes in this edition:

- *Scraping the barnacles.* tl;dr. For those not familiar with Internet slang, this abbreviation means "too long, didn't read." Sadly, many students take this approach to textbooks. My main goal in this revision, therefore, has been a renewed commitment to brevity. In particular, I took up the task of scraping off the barnacles that have accumulated over many editions. More important than what has been added to this edition is what has been taken out. This task has benefited from surveys of instructors who use the book. I have kept what most instructors consider essential and taken out what most consider superfluous.
- *Streaming coverage of consumption and investment.* The material on the microeconomic foundations of consumption and investment has been condensed into a single, more accessible chapter.
- *New topic in Chapter 9*. The role of culture in economic growth.
- *New topic in Chapter 12*. The curious case of negative interest rates.
- *New topic in Chapter 18*. The stress tests that regulators are using to evaluate banks' safety and soundness.
- *New assessment tool.* This edition includes a new pedagogical feature. Every chapter concludes with a Quick Quiz of six multiple-choice questions. Students can use these quizzes to immediately test their understanding of the core concepts in the chapter. The quiz answers are available at the end of each chapter.
- *Updated data.* As always, the book has been fully updated. All the data are as current as possible.

Despite these changes, my goal remains the same as in previous editions: to offer students the clearest, most up-to-date, most accessible course in macroeconomics in the fewest words possible.

The Arrangement of Topics

My strategy for teaching macroeconomics is first to examine the long run, when prices are flexible, and then to examine the short run, when prices are sticky. This approach has several advantages. First, because the classical dichotomy permits the separation of real and monetary issues, the long-run material is easier for students. Second, when students begin studying short-run fluctuations, they understand the long-run equilibrium around which the economy is fluctuating. Third, beginning with market-clearing models clarifies the link between macroeconomics and microeconomics. Fourth, students learn first the material that is less controversial. For all these reasons, the strategy of beginning with long-run classical models simplifies the teaching of macroeconomics.

Let’s now move from strategy to tactics. What follows is a whirlwind tour of the book.

Part One, Introduction

The introductory material in Part One is brief so that students can get to the core topics quickly. Chapter 1 discusses the questions that macroeconomists address and the economist’s approach of building models to explain the world. Chapter 2 introduces the data of macroeconomics, emphasizing gross domestic product, the consumer price index, and the unemployment rate.

Part Two, Classical Theory: The Economy in the Long Run

Part Two examines the long run, over which prices are flexible. Chapter 3 presents the classical model of national income. In this model, the factors of production and the production technology determine the level of income, and the marginal products of the factors determine its distribution to households. In addition, the model shows how fiscal policy influences the allocation of the economy’s resources among consumption, investment, and government purchases, and it highlights how the real interest rate equilibrates the supply and demand for goods and services.

Money and the price level are introduced next. Chapter 4 examines the monetary system and the tools of monetary policy. Chapter 5 begins the discussion of the effects of monetary policy. Because prices are assumed to be flexible, the chapter presents the ideas of classical monetary theory: the quantity theory of money, the inflation tax, the Fisher effect, the social costs of inflation, and the causes and costs of hyperinflation.

The study of open-economy macroeconomics begins in Chapter 6. Maintaining the assumption of full employment, this chapter presents models that explain the trade balance and the exchange rate. Various policy issues are addressed: the relationship between the budget deficit and the trade deficit, the macroeconomic impact of protectionist trade policies, and the effect of monetary policy on the value of a currency in the market for foreign exchange.

Chapter 7 relaxes the assumption of full employment, discussing the dynamics of the labor market and the natural rate of unemployment. It examines various causes of unemployment, including job search, minimum-wage laws, union power, and efficiency wages. It also presents some important facts about patterns of unemployment.

Part Three, Growth Theory: The Economy in the Very Long Run

Part Three makes the classical analysis of the economy dynamic with the tools of growth theory. Chapter 8 introduces the Solow growth model, emphasizing capital accumulation and population growth. Chapter 9 then adds technological progress to the Solow model. It uses the model to discuss growth experiences around the world as well as public policies that influence the level and growth of the standard of living. Chapter 9 also introduces students to the modern theories of endogenous growth.

Part Four, Business Cycle Theory: The Economy in the Short Run

Part Four examines the short run, when prices are sticky. It begins in Chapter 10 by examining the key facts that describe short-run fluctuations in economic activity. The chapter then introduces the model of aggregate supply and aggregate demand, as well as the role of stabilization policy. Subsequent chapters refine the ideas introduced in this chapter.

Chapters 11 and 12 look more closely at aggregate demand. Chapter 11 presents the Keynesian cross and the theory of liquidity preference and uses these models as building blocks for the *IS–LM* model. Chapter 12 uses the *IS–LM* model to explain economic fluctuations and the aggregate demand curve, concluding with an extended case study of the Great Depression.

The discussion of short-run fluctuations continues in Chapter 13, which focuses on aggregate demand in an open economy. This chapter presents the Mundell–Fleming model and shows how monetary and fiscal policies affect the economy under floating and fixed exchange-rate systems. It also discusses the question of whether exchange rates should be floating or fixed.

Chapter 14 looks more closely at aggregate supply. It examines various approaches to explaining the short-run aggregate supply curve and discusses the short-run tradeoff between inflation and unemployment.

Part Five, Topics in Macroeconomic Theory and Policy

Once students have a solid command of standard models, the book offers them various optional chapters that dive deeper into macroeconomic theory and policy.

Chapter 15 develops a dynamic model of aggregate demand and aggregate supply. It builds on ideas that students have already encountered and uses those ideas as stepping-stones to take students closer to the frontier of knowledge about short-run fluctuations. The model presented here is a simplified version of modern dynamic, stochastic, general equilibrium (DSGE) models.

Chapter 16 considers the debate over how policymakers should respond to short-run fluctuations. It emphasizes two questions: Should monetary and fiscal policy be active or passive? Should policy be conducted by rule or discretion? The chapter presents arguments on both sides of these questions.

Chapter 17 focuses on debates over government debt and budget deficits. It gives a broad picture of the magnitude of government indebtedness, discusses why measuring budget deficits is not always straightforward, recaps the traditional view of the effects of government debt, presents Ricardian equivalence as an alternative view, and examines various other perspectives on government debt. As in the previous chapter, students are not handed conclusions but are given tools to evaluate alternative viewpoints on their own.

Chapter 18 discusses the financial system and its linkages to the overall economy. It begins by examining what the financial system does: financing investment, sharing risk, dealing with asymmetric information, and fostering growth. It then discusses the causes of financial crises, their macroeconomic impact, and the policies that might mitigate their effects and reduce their likelihood.

Chapter 19 analyzes some of the microeconomics behind consumption and investment decisions. It discusses various theories of consumer behavior, including the Keynesian consumption function, Modigliani’s life-cycle hypothesis, Friedman’s permanent-income hypothesis, Hall’s random-walk hypothesis, and Laibson’s model of instant gratification. It also examines the theory behind the investment function, focusing on business fixed investment and including topics such as the cost of capital, Tobin’s q, and the role of financing constraints.

Epilogue

The book ends with an epilogue that reviews the broad lessons about which most macroeconomists agree and some important open questions. Regardless of which chapters an instructor covers, this capstone chapter can be used to remind students how the many models and themes of macroeconomics relate to one another. Here and throughout the book, I emphasize that despite the disagreements among macroeconomists, there is much that we know about how the economy works.

Alternative Routes Through the Text

Instructors of intermediate macroeconomics have different preferences about the choice and organization of topics. I kept this in mind while writing the book so that it would offer a degree of flexibility. Here are a few ways that instructors might consider rearranging the material:

- Some instructors are eager to cover short-run economic fluctuations. For such a course, I recommend covering Chapters 1 through 5 so that students are grounded in the basics of classical theory and then jumping to Chapters 10, 11, 12, and 14 to cover the model of aggregate demand and aggregate supply.
- Some instructors are eager to cover long-run economic growth. These instructors can cover Chapters 8 and 9 immediately after Chapter 3.
- An instructor who wants to defer (or even skip) open-economy macroeconomics can put off Chapters 6 and 13 without loss of continuity.
- An instructor who wants to emphasize macroeconomic policy can skip Chapters 8, 9, and 15 in order to get to Chapters 16, 17, and 18 more quickly.
- An instructor who wants to stress the microeconomic foundations of macroeconomics can cover Chapter 19 early in the course, even after Chapter 3.

The successful experiences of hundreds of instructors with previous editions suggest this text nicely complements a variety of approaches to the field.

Learning Tools

I am pleased that students have found the previous editions of this book user-friendly. I have tried to make this tenth edition even more so.

Case Studies

Economics comes to life when it is applied to understanding actual events. Therefore, the numerous case studies are an important learning tool, integrated closely with the theoretical material presented in each chapter. The frequency with which these case studies occur ensures that a student does not have to grapple with an overdose of theory before seeing the theory applied. Students report that the case studies are their favorite part of the book.

FYI Boxes

These boxes present ancillary material “for your information.” I use these boxes to clarify difficult concepts, to provide additional information about the tools of economics, and to show how economics impacts our daily lives.

Graphs

Understanding graphical analysis is a central part of learning macroeconomics, and I have worked hard to make the figures easy to follow. I often use comment boxes within figures to describe and draw attention to the key points that the figures illustrate. The pedagogical use of color, detailed captions, and comment boxes makes it easier for students to learn and review the material.

Mathematical Notes

I use occasional mathematical footnotes to keep difficult material out of the body of the text. These notes make an argument more rigorous or present a proof of a mathematical result. They can be skipped by students who have not been introduced to the necessary mathematical tools.

Quick Quizzes

Every chapter ends with six multiple-choice questions, which students can use to test themselves on what they have just read. The answers are provided at the end of each chapter. These quizzes are new to this edition.

Chapter Summaries

Every chapter includes a brief, nontechnical summary of its major lessons. Students can use the summaries to place the material in perspective and to review for exams.

Key Concepts

Learning the language of a field is a major part of any course. Within the chapter, each key concept is in **boldface** when it is introduced. At the end of the chapter, the key concepts are listed for review.

Questions for Review

Students are asked to test their understanding of a chapter's basic lessons in the Questions for Review.

Problems and Applications

Every chapter includes Problems and Applications designed for homework assignments. Some are numerical applications of the theory in the chapter. Others encourage students to go beyond the material in the chapter by addressing new issues that are closely related to the chapter topics. In most of the core chapters, a few problems are identified with this icon: Work It Out. For each of these problems, students can find a Work It Out tutorial on SaplingPlus for *Macroeconomics*, 10e: **https://macmillanlearning.com/sapling**.

Chapter Appendices

Several chapters include appendices that offer additional material, sometimes at a higher level of mathematical sophistication. These appendices are designed so that instructors can cover certain topics in greater depth if they wish. The appendices can be skipped altogether without loss of continuity.

Glossary

To help students become familiar with the language of macroeconomics, a glossary of more than 250 terms is provided at the back of the book.

International Editions

The English-language version of this book has been used in dozens of countries. To make the book more accessible for students around the world, editions are (or will soon be) available in 17 other languages: Armenian, Chinese (Simplified and Complex), French, German, Greek, Hungarian, Indonesian, Italian, Japanese, Korean, Portuguese, Romanian, Russian, Spanish, Ukrainian, and Vietnamese. In addition, a Canadian adaptation coauthored with William Scarth (McMaster University) and a European adaptation coauthored with Mark Taylor (University of Warwick) are available. Instructors who would like information about these versions of the book should contact Worth Publishers.

Acknowledgments

Since I started writing the first edition of this book, I have benefited from the input of many reviewers and colleagues in the economics profession. Now that the book is in its tenth edition, these people are too numerous to list in their entirety. However, I continue to be grateful for their willingness to have given up their scarce time to help me improve the economics and pedagogy of this text. Their advice has made this book a better teaching tool for hundreds of thousands of students around the world.

I would like to mention the instructors whose recent input shaped this new edition:

David Aadland
University of Wyoming
Lian An
University of Northern Florida
Samuel K. Andoh
Southern Connecticut State University
Dennis Avola
Northeastern University
Mustapha Ibn Boamah
University of New Brunswick, Saint John
Jeffrey Buser
The Ohio State University
Kenneth I. Carlaw
University of British Columbia, Okanagan
Sel Dibooglu
University of Missouri, St. Louis
Oguzhan Dincer
Illinois State University
Christi-Anna Durodola
University of Winnipeg
Per Fredriksson
University of Louisville
Mark J. Gibson
Washington State University
Lisa R. Gladson
St. Louis University
David W. Johnson
University of Wisconsin–Madison
J.B. Kim
Oklahoma State University
Ryan Lee
Indiana University
Meghan Millea
Mississippi State University
Daniel Moncayo
Brigham Young University

Robert Murphy
Boston College
John Neri
University of Maryland
Russell M. Price
Howard University
Raul Razo-Garcia
Carleton University
Subarna Samanta
The College of New Jersey
Ruben Sargsyan
California State University, Chico
Fahlino F. Sjuib
Framingham State University
Julie K. Smith
Lafayette College
Peter Summers
High Point University
Ariuntungalag Taivan
University of Minnesota–Duluth
Kiril Tochkov
Texas Christian University
Christian vom Lehn
Brigham Young University
Paul Wachtel
New York University

In addition, I am grateful to Nina Vendhan, a student at Harvard, who helped me update the data and refine my prose. Nina, along with my son Nick Mankiw, also helped me proofread the book.

The people at Worth Publishers have continued to be congenial and dedicated. I would like to thank Catherine Woods, Vice President, Content Management; Charles Linsmeier, Senior Vice President, Content Strategy; Shani Fisher, Director of Content and Assessment; Simon Glick, Executive Program Manager; Andrew Zierman, Marketing Manager; Travis Long, Learning Solutions Specialist; Noel Hohnstine, Director of Media Editorial; Nikolas Toner, Associate Media Editor; Assessment Manager, Kristyn Brown; Joshua Hill, Assessment Editor; Lukia Kliossis, Development Editor; Courtney Lindwall, Assistant Editor; Amanda Gaglione, Editorial Assistant; Hannah Aronowitz, Editorial Intern; Lisa Kinne, Senior Managing Editor; Tracey Kuehn, Director, Content Management Enhancement; Diana Blume, Director of Design, Content Management; and Kitty Wilson, Copyeditor.

Many other people have made valuable contributions as well. Most important, Jane Tufts, freelance

developmental editor, worked her magic on this book once again, confirming that she's the best in the business. Alexandra Nickerson did a great job preparing the index. Deborah Mankiw, my wife and in-house editor, continued to be the first reader of new material, providing the right mix of criticism and encouragement.

Finally, I would like to thank my three children, Catherine, Nicholas, and Peter. They helped immensely with this revision—both by providing a pleasant distraction and by reminding me that textbooks are written for the next generation.

May 2018

Contents

The Science of Macroeconomics

Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

> The whole of science is nothing more than a refinement of everyday thinking.
>
> —Albert Einstein

When Albert Einstein made the above observation, he was probably referring to the natural sciences like physics and chemistry. But the statement also applies to social sciences such as economics. As a participant in the economy, and as a citizen in a democracy, you cannot help but think about economic issues as you go about your life or when you enter the voting booth. But if you are like most people, your everyday thinking about economics has been casual rather than rigorous (or at least it was before you took your first economics course). The goal of studying economics is to refine that thinking. This book aims to help you in that endeavor, focusing on the part of the field called **macroeconomics**, which studies the forces that influence the economy as a whole.

1-1 What Macroeconomists Study

Why have some countries experienced rapid growth in incomes over the past century while others have stayed mired in poverty? Why do some countries have high rates of inflation while others maintain stable prices? Why do all countries experience recessions and depressions — recurrent periods of falling incomes and rising unemployment — and how can government policy reduce the frequency and severity of these episodes? Macroeconomics attempts to answer these and many related questions.

To appreciate the importance of macroeconomics, you need only visit a news website. Every day you can see headlines such as INCOME GROWTH REBOUNDS, FED MOVES TO COMBAT INFLATION, or JOBS REPORT SENDS STOCKS LOWER. These macroeconomic events may seem abstract, but they touch all of our lives. Business executives forecasting the demand for their products must guess how fast consumers' incomes will grow. Senior citizens living on fixed incomes wonder how quickly prices will rise. Recent college graduates looking for employment hope that the economy will boom and that firms will be hiring.

Because the state of the economy affects everyone, macroeconomic issues play a central role in national political debates. Voters are aware of how the economy is doing, and they know that government policy can affect the economy in powerful ways. As a result, the popularity of an incumbent president often rises when the economy is doing well and falls when it is doing poorly.

Macroeconomic issues are also central to world politics, and the international news is filled with macroeconomic questions. Was it a good move for much of Europe to adopt a common currency? Should China maintain a fixed exchange rate against the U.S. dollar? Why is the United States running large trade deficits? How can poor nations raise their standards of living? When world leaders meet, these topics are often high on the agenda.

Although the job of making economic policy belongs to world leaders, the job of explaining the workings of the economy as a whole falls to macroeconomists. To this end, macroeconomists collect data on incomes, prices, unemployment, and many other variables from different time periods and different countries. They then attempt to formulate general theories to explain these data. Like astronomers studying the evolution of stars or biologists studying the evolution of species, macroeconomists usually cannot conduct controlled experiments in a laboratory. Instead, they must make use of the data that history gives them. Macroeconomists observe that economies differ across countries and that they change over time. These observations provide both the motivation for developing macroeconomic theories and the data for testing them.

To be sure, macroeconomics is an imperfect science. The macroeconomist's ability to predict the future course of economic events is no better than the meteorologist's ability to predict next month's weather. But, as you will see, macroeconomists know quite a lot about how economies work. This knowledge is useful both for explaining economic events and for formulating economic policy.

Every era has its own economic problems. In the 1970s, Presidents Richard Nixon, Gerald Ford, and Jimmy Carter all wrestled in vain with a rising inflation rate. In the 1980s, inflation subsided, but Presidents Ronald Reagan and George H. W. Bush presided over large federal budget deficits. In the 1990s, with President Bill Clinton in the Oval Office, the economy and stock market enjoyed a remarkable boom, and the federal budget turned from deficit to surplus. As Clinton left office, however, the stock market was in retreat, and the economy was heading into recession. In 2001, President George W. Bush reduced taxes to help end the recession, but the tax cuts contributed to a reemergence of budget deficits.

President Barack Obama moved into the White House in 2009 during a period of heightened economic turbulence. The economy was reeling from a financial crisis driven by falling housing prices, rising mortgage defaults, and the bankruptcy or near-bankruptcy of many large and economically significant financial institutions. As the crisis spread, it raised the specter of the Great Depression of the 1930s, when in its worst year one out of four Americans who wanted to work could not find a job. In 2008 and 2009, officials in the Treasury, Federal Reserve, and other parts of government acted vigorously to prevent a recurrence of that outcome.

In some ways, policymakers succeeded; the unemployment rate peaked at 10 percent in 2009. But the downturn was nonetheless severe, and the subsequent recovery was slow. Total income in the economy, adjusted for inflation, grew at an average rate of 1.3 percent per year from 2006 to 2016, well below the historical norm of 3.2 percent per year. These events helped set the stage for President Donald Trump's campaign slogan of 2016: "Make America Great Again."

Macroeconomic history is not a simple story, but it provides a rich motivation for macroeconomic theory. While the basic principles of macroeconomics do not change from decade to decade, the macroeconomist must apply these principles with flexibility and creativity to meet changing circumstances.

CASE STUDY

The Historical Performance of the U.S. Economy

Economists use many types of data to measure the performance of an economy. Three macroeconomic variables are especially important: real gross domestic product (GDP), the inflation rate, and the unemployment rate. **Real GDP** measures the total income of everyone in the economy (adjusted for the level of prices). The **inflation rate** measures how fast prices are rising. The **unemployment rate** measures the fraction of the labor force that is out of work. Macroeconomists study how these variables are determined, why they change over time, and how they interact with one another.

Figure 1-1 shows real GDP per person in the United States. Two aspects of this figure are noteworthy. First, real GDP grows over time. Real GDP per person today is more than eight times higher than it was in 1900. This growth in average income allows us to enjoy a much higher standard of living than our great-grandparents did.

Second, although real GDP rises in most years, this growth is not steady. There are repeated periods during which real GDP falls, the most dramatic instance being the early 1930s. Such periods are called **recessions** if they are mild and **depressions** if they are more severe. Not surprisingly, periods of declining income are associated with substantial economic hardship.

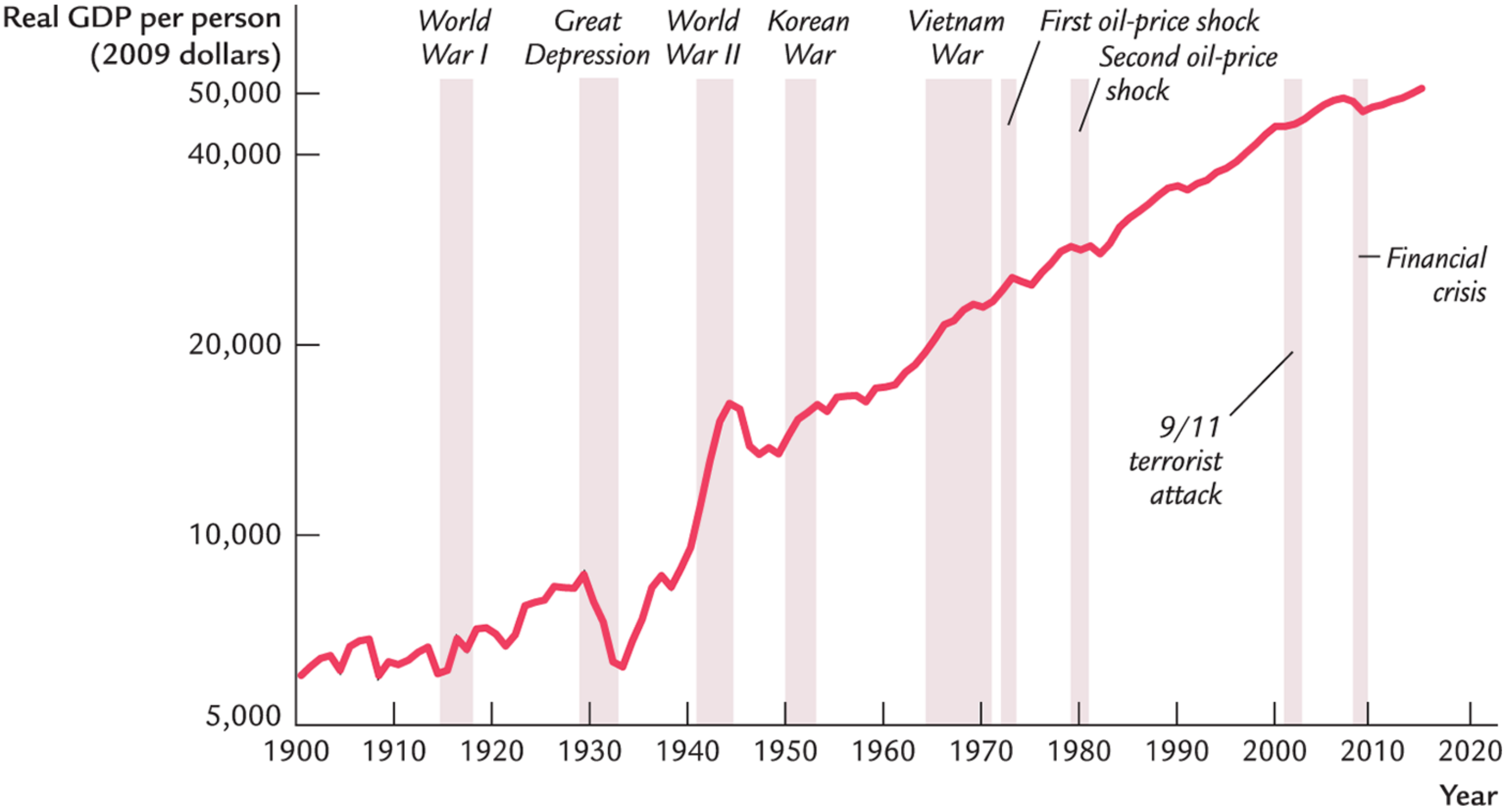


FIGURE 1-1 Real GDP Per Person in the U.S. Economy Real GDP measures the total income of everyone in the economy, and real GDP per person measures the income of the average person in the economy. This figure shows that real GDP per person tends to grow over time and that this normal growth is sometimes interrupted by periods of declining income, called recessions or depressions.

Note: Real GDP is plotted here on a logarithmic scale. On such a scale, equal distances on the vertical axis represent equal *percentage* changes. Thus, the distance between $5,000 and $10,000 (a 100 percent change) is the same as the distance between $10,000 and $20,000 (a 100 percent change).

Data from: U.S. Department of Commerce, Economic History Association.

Figure 1-2 shows the U.S. inflation rate. You can see that inflation varies substantially over time. In the first half of the twentieth century, the inflation rate averaged only slightly above zero. Periods of falling prices, called **deflation**, were almost as common as periods of rising prices. By contrast, inflation has been the norm during the past half century. Inflation became most severe during the late 1970s, when prices rose at a rate of almost 10 percent per year. In recent years, the inflation rate has been about 2 percent per year, indicating that prices have been fairly stable.

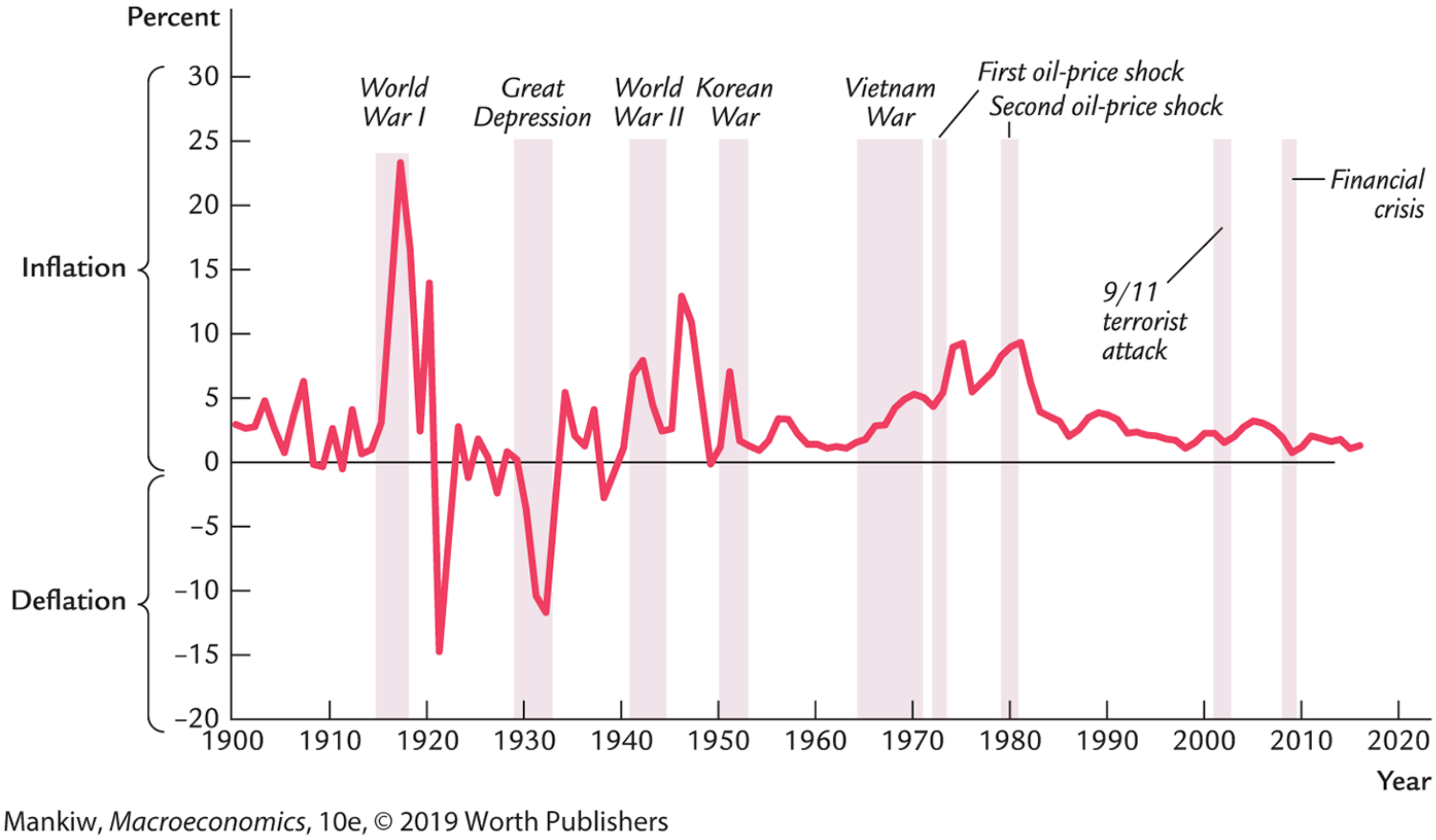
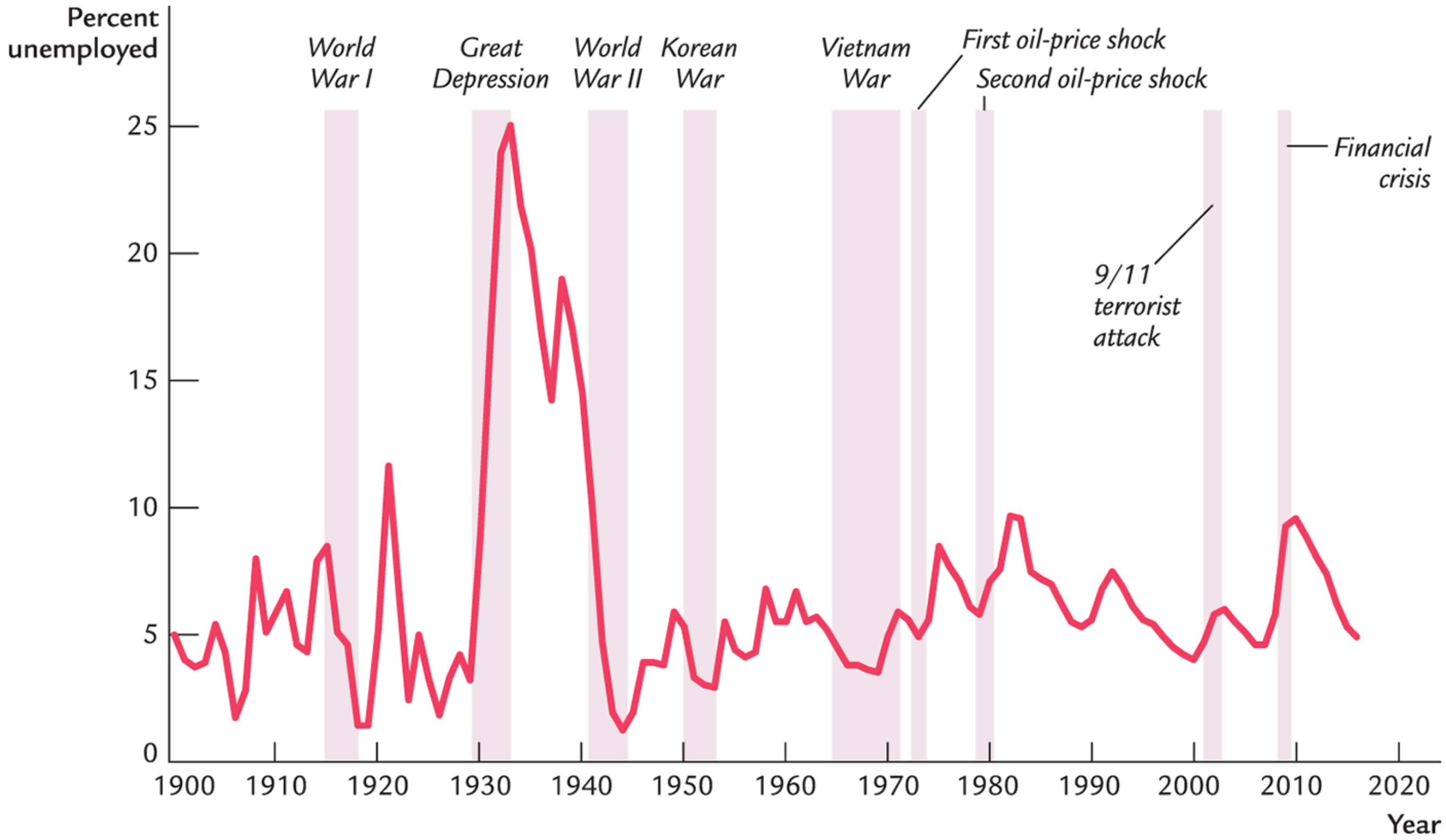


FIGURE 1-2 The Inflation Rate in the U.S. Economy The inflation rate measures the percentage change in the average level of prices from the year before. When the inflation rate is above zero, prices are rising. When it is below zero, prices are falling. If the inflation rate declines but remains positive, prices are rising but at a slower rate.

Note: The inflation rate is measured here using the GDP deflator.

Data from: U.S. Department of Commerce, Economic History Association.

Figure 1-3 shows the U.S. unemployment rate. Notice that there is always some unemployment in the economy. In addition, although the unemployment rate has no long-term trend, it varies substantially from year to year. Recessions and depressions are associated with unusually high unemployment. The highest rates of unemployment were reached during the Great Depression of the 1930s. The worst economic downturn since the Great Depression occurred in the aftermath of the financial crisis of 2008–2009, when unemployment rose substantially. Even several years after the crisis (called the “Great Recession”), unemployment remained high. Unemployment did not return to its 2007 level until 2016.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 1-3 The Unemployment Rate in the U.S. Economy The unemployment rate measures the percentage of people in the labor force who do not have jobs. This figure shows that the economy always has some unemployment and that the amount fluctuates from year to year.

Data from: U.S. Department of Labor, U.S. Census Bureau.

These three figures offer a glimpse at the history of the U.S. economy. In the chapters that follow, we first discuss how these variables are measured and then develop theories to explain how they behave. ■

1-2 How Economists Think

Economists often study politically charged issues, but they try to address these issues with a scientist's objectivity. Like any science, economics has its own set of tools — terminology, data, and a way of thinking — that can seem foreign and arcane to the layperson. The best way to become familiar with these tools is to practice using them, and this book affords you ample opportunity to do so. To make these tools less forbidding, however, let's discuss a few of them here.

Theory as Model Building

Children learn about the world by playing with toy versions of real objects. For instance, they often put together models of cars, trains, or planes. These models are not realistic, but the model-builder learns a lot from them nonetheless. The model illustrates the essence of the object it is designed to resemble. (In addition, for many children, building models is fun.)

Economists also use **models** to understand the world, but an economist's model is more likely to be made of symbols and equations than plastic and glue. Economists build their "toy economies" to explain economic variables, such as GDP, inflation, and unemployment. Economic models illustrate, often in mathematical terms, the relationships among the variables. Models are useful because they help us dispense with irrelevant details and focus on underlying connections. (In addition, for many economists, building models is fun.)

Models have two kinds of variables: endogenous variables and exogenous variables. **Endogenous variables** are those variables that a model explains. **Exogenous variables** are those variables that a model takes as given. The purpose of a model is to show how the exogenous variables influence the endogenous variables. In other words, as Figure 1-4 illustrates, exogenous variables come from outside the model and serve as the model's input, whereas endogenous variables are determined within the model and are the model's output.

FIGURE 1-4 How Models Work Models are simplified theories that show the key relationships among economic variables. The exogenous variables are those that come from outside the model. The endogenous variables are those that the model explains. The model shows how changes in the exogenous variables affect the endogenous variables.

To make these ideas more concrete, let's review the most celebrated of all economic models — the model of supply and demand. Imagine that an economist wants to figure out what factors influence the price of pizza and the quantity of pizza sold. She would develop a model to describe the behavior of pizza buyers, the behavior of pizza sellers, and their interaction in the market for pizza. For example, the economist supposes that the quantity of pizza demanded by consumers Q^d depends on the price of pizza P and on aggregate income Y. This relationship is expressed in the equation

$$Q^d = D(P, Y),$$

where $D(\,)$ represents the demand function. Similarly, the economist supposes that the quantity of pizza supplied by pizzerias Q^s depends on the price of pizza P and on the price of materials P_m, such as cheese, tomatoes, flour, and anchovies. This relationship is expressed as

$$Q^s = S(P, P_m),$$

where $S(\,)$ represents the supply function. Finally, the economist assumes that the price of pizza adjusts to bring the quantity supplied and quantity demanded into balance:

$$Q^s = Q^d.$$

These three equations compose a model of the market for pizza.

The economist illustrates the model with a supply-and-demand diagram, as in Figure 1-5. The demand curve shows the relationship between the quantity of pizza demanded and the price of pizza, holding aggregate income constant. The demand curve slopes downward because a higher price of pizza encourages consumers to buy less pizza and switch to, say, hamburgers and tacos. The supply curve shows the relationship between the quantity of pizza supplied and the price of pizza, holding the price of materials constant. The supply curve slopes upward because a higher price of pizza makes selling pizza more profitable, which encourages pizzerias to produce more of it. The equilibrium for the market is the price and quantity at which the supply and demand curves intersect. At the equilibrium price, consumers choose to buy the amount of pizza that pizzerias choose to produce.

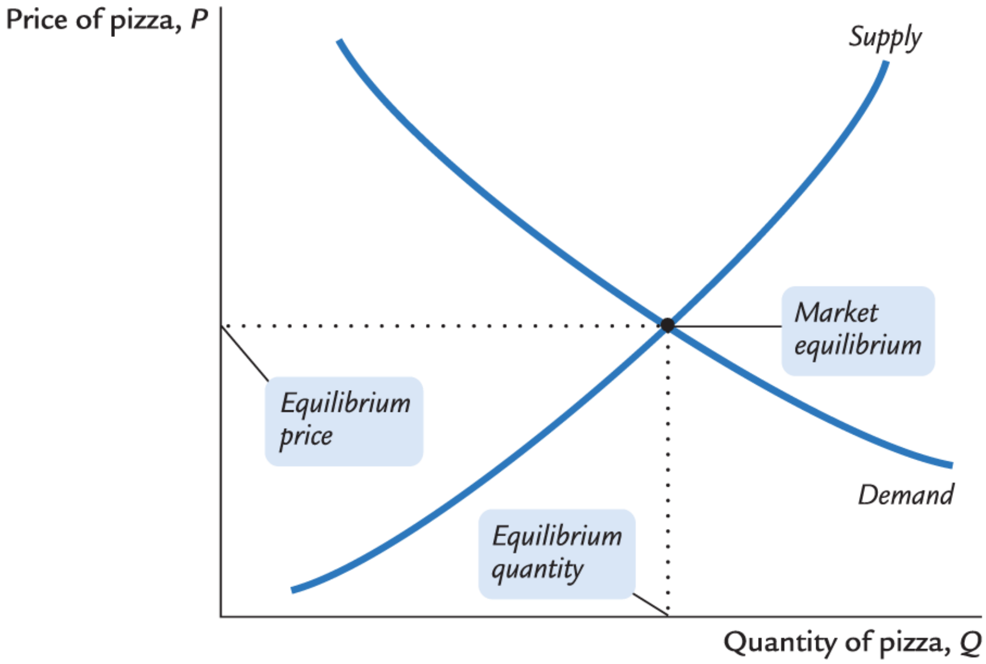


FIGURE 1-5 The Model of Supply and Demand The most famous economic model is that of supply and demand for a good or service — in this case, pizza. The demand curve is a downward-sloping curve relating the price of pizza to the quantity of pizza that consumers demand. The supply curve is an upward-sloping curve relating the price of pizza to the quantity of pizza that pizzerias supply. The price of pizza adjusts until the quantity supplied equals the quantity demanded. The point where the two curves cross is the market equilibrium, which shows the equilibrium price of pizza and the equilibrium quantity of pizza.

This model of the pizza market has two exogenous variables and two endogenous variables. The exogenous variables are aggregate income and the price of materials. The model does not explain them but instead takes them as given (perhaps to be explained by another model). The endogenous variables are the price of pizza and the quantity of pizza exchanged. These are the variables that the model explains.

The model can be used to show how a change in any exogenous variable affects both endogenous variables. For example, if aggregate income increases, then the demand for pizza increases, as in panel (a) of Figure 1-6. The model shows that both the equilibrium price and the equilibrium quantity of pizza rise. Similarly, if the price of materials increases, then the supply of pizza decreases, as in panel (b) of Figure 1-6. The model shows that in this case, the equilibrium price of pizza rises, while the equilibrium quantity of pizza falls. Thus, the model shows how changes either in aggregate income or in the price of materials affect price and quantity in the market for pizza.

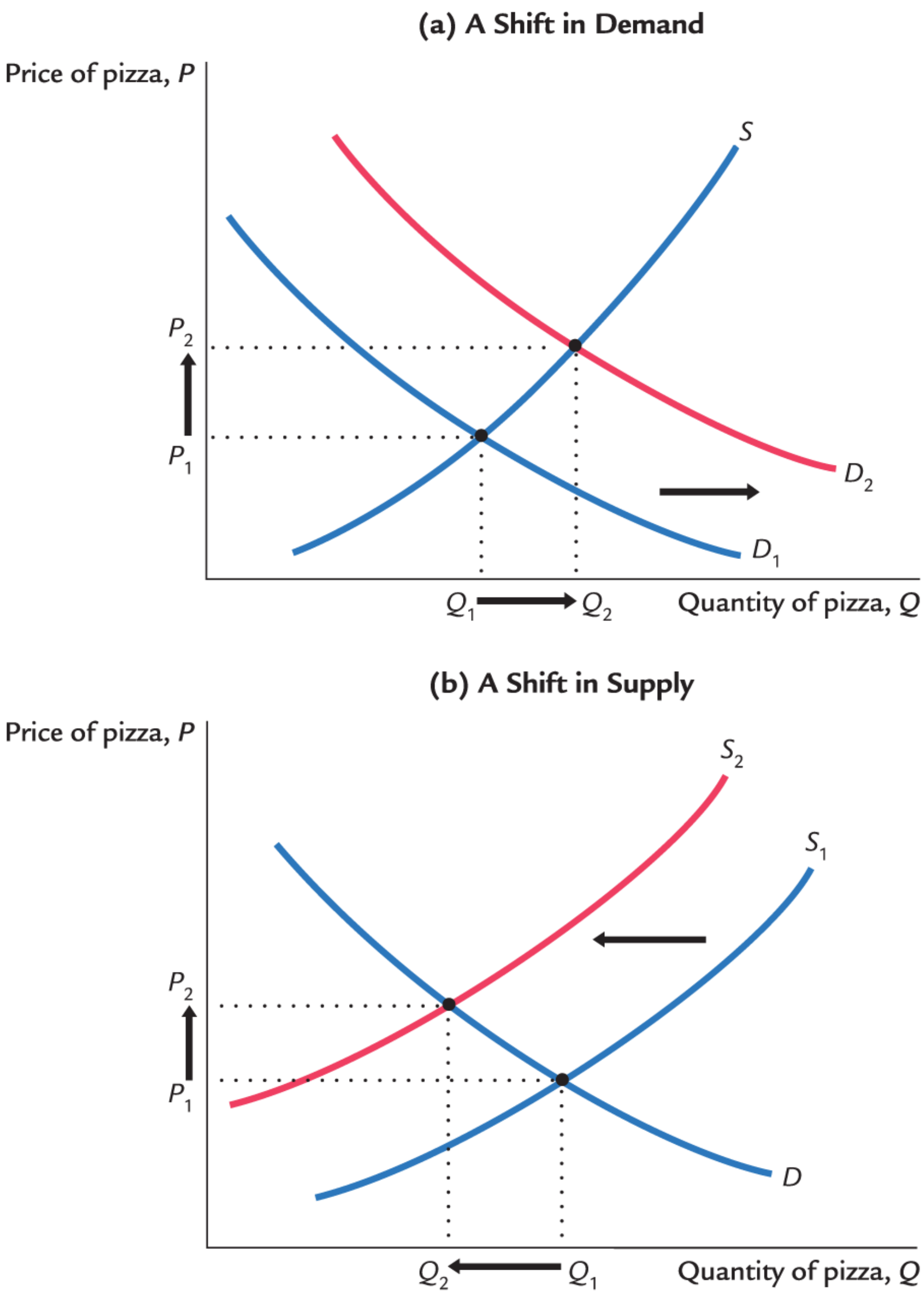


FIGURE 1-6 Changes in Equilibrium In panel (a), a rise in aggregate income causes the demand for pizza to increase: at any given price, consumers now want to buy more pizza. This is represented by a rightward shift in the demand curve from D1D_1 to D2.D_2. The market moves to the new intersection of supply and demand. The equilibrium price rises from P1P_1 to P2,P_2, and the equilibrium quantity of pizza rises from Q1Q_1 to Q2.Q_2. In panel (b), a rise in the price of materials decreases the supply of pizza: at any given price, pizzerias find that the sale of pizza is less profitable and therefore choose to produce less pizza. This is represented by a leftward shift in the supply curve from S1S_1 to S2.S_2. The market moves to the new intersection of supply and demand. The equilibrium price rises from P1P_1 to P2,P_2, and the equilibrium quantity falls from Q1Q_1 to Q2.Q_2.

Like all models, this model of the pizza market makes simplifying assumptions. The model does not take into account, for example, that every pizzeria is in a different location. For each customer, one pizzeria is more convenient than the others, and thus pizzerias have some ability to set their own prices. The model assumes that there is a single price for pizza, but in fact there could be a different price at each pizzeria.

How should we react to the model's lack of realism? Should we discard the simple model of pizza supply and demand? Should we attempt to build a more complex model with diverse pizza prices? The answers to these questions depend on our purpose. If our goal is to explain how the price of cheese affects the average price of pizza and the amount of pizza sold, then the diversity of pizza prices is probably not important. The simple model of the pizza market does a good job of addressing that issue. Yet if our goal is to explain why towns with ten pizzerias have lower pizza prices than towns with only two, the simple model is less useful.

The art in economics lies in judging when a simplifying assumption (such as assuming a single price of pizza) clarifies our thinking and when it misleads us. Simplification is a necessary part of building a useful model: any model constructed to be completely realistic would be too complicated for anyone to understand. Yet if models assume away features of the economy that are crucial to the issue at hand, they may lead us to wrong conclusions. Economic modeling therefore requires care and common sense.

FYI

Using Functions to Express Relationships Among Variables

All economic models express relationships among economic variables. Often, these relationships are expressed as functions. A *function* is a mathematical concept that shows how one variable depends on a set of other variables. For example, in the model of the pizza market, we said that the quantity of pizza demanded depends on the price of pizza and on aggregate income. To express this, we use functional notation to write

$$Q^d = D(P, Y).$$

This equation says that the quantity of pizza demanded Q^d is a function of the price of pizza P and aggregate income Y. In functional notation, the variable preceding the parentheses denotes the function. In this case, $D(\,)$ is the function expressing how the variables in parentheses determine the quantity of pizza demanded.

If we knew more about the pizza market, we could give a numerical formula for the quantity of pizza demanded. For example, we might be able to write

$$Q^d = 60 - 10P + 2Y.$$

In this case, the demand function is

$$D(P, Y) = 60 - 10P + 2Y.$$

For any price of pizza and aggregate income, this function gives the corresponding quantity of pizza demanded.

For example, if aggregate income is $10 and the price of pizza is $2, then the quantity of pizza demanded is 60 pies; if the price of pizza rises to $3, the quantity of pizza demanded falls to 50 pies.

Functional notation allows us to express the idea that variables are related, even when we do not have enough information to indicate the precise numerical relationship. For example, we might know that the quantity of pizza demanded falls when the price rises from $2 to $3, but we might not know by how much it falls. In this case, functional notation is useful: as long as we know that a relationship among the variables exists, we can express that relationship using functional notation.

The Use of Multiple Models

Macroeconomists study many facets of the economy. For example, they examine the role of saving in economic growth, the impact of minimum-wage laws on unemployment, the effect of inflation on interest rates, and the influence of trade policy on the trade balance and exchange rate.

Economists use models to address all of these issues, but no single model can answer every question. Just as carpenters use different tools for different tasks, economists use different models to explain different phenomena. Students of macroeconomics must keep in mind that there is no single “correct” model that applies to every economic question. Instead, there are many models, each of which is useful for shedding light on a particular facet of the economy. The field of macroeconomics is like a Swiss Army knife — a set of complementary but distinct tools that can be applied in different ways in different circumstances.

This book presents many different models that address different questions and make different assumptions. Remember that a model is only as good as its assumptions and that an assumption that is useful for some purposes may be misleading for others. When using a model, the economist must keep in mind the underlying assumptions and judge whether they are reasonable for studying the matter at hand.

Prices: Flexible Versus Sticky

Throughout this book, one group of assumptions will prove especially important: those concerning the speed at which wages and prices adjust to changing conditions. Economists normally presume that the price of a good or a service moves quickly to bring quantity supplied and quantity demanded into balance. In other words, they assume that markets are normally in equilibrium, so the price of any good or service is found where the supply and demand curves intersect. This assumption, called **market clearing**, is central to the model of the pizza market discussed earlier. For answering most questions, economists use market-clearing models.

However, the assumption of *continuous* market clearing is not entirely realistic. For markets to clear continuously, prices must adjust instantly to changes in supply and demand. In fact, many wages and prices adjust slowly. Labor contracts often set wages for up to three years. Many firms leave their product prices the same for long periods of time; for example, magazine publishers change their newsstand prices only every three or four years. Although market-clearing models assume that all wages and prices are **flexible**, in the real world some wages and prices are **sticky**.

The apparent stickiness of prices does not make market-clearing models useless. After all, prices are not stuck forever; eventually, they adjust to changes in supply and demand. Market-clearing models might not describe the economy at every instant, but they do describe the equilibrium toward which the economy gravitates. Therefore, most macroeconomists believe that price flexibility is a good assumption for studying long-run issues, such as the growth in real GDP that we observe from decade to decade.

For studying short-run issues, such as year-to-year fluctuations in real GDP and unemployment, the assumption of price flexibility is less plausible. Over short periods, many prices in the economy are fixed at predetermined levels. Therefore, most macroeconomists believe that price stickiness is a better assumption for studying the short-run behavior of the economy.

Microeconomic Thinking and Macroeconomic Models

Microeconomics is the study of how households and firms make decisions and how these decisionmakers interact in the marketplace. A central principle of microeconomics is that households and firms *optimize* — they do the best they can for themselves, given their objectives and the constraints they face. In microeconomic models, households choose their purchases to maximize their level of satisfaction, called *utility*, and firms make production decisions to maximize their profits.

Because economy-wide events arise from the interaction of many households and firms, macroeconomics and microeconomics are inextricably linked. When we study the economy as a whole, we must consider the decisions of individual economic actors. For example, to understand what determines total consumer spending, we must think about a family deciding how much to spend today and how much to save for the future. To understand what determines total investment spending, we must think about a firm deciding whether to build a new factory. Because aggregate variables are the sum of the variables describing many individual decisions, macroeconomic theory rests on a microeconomic foundation.

Although microeconomic decisions underlie all economic models, in many models the optimizing behavior of households and firms is implicit rather than explicit. The model of the pizza market we discussed earlier is an example. Households' decisions about how much pizza to buy underlie the demand for pizza, and pizzerias' decisions about how much pizza to produce underlie the supply of pizza. Presumably, households make their decisions to maximize utility, and pizzerias make their decisions to maximize profit. Yet the model does not focus on how these microeconomic decisions are made; instead, it leaves these decisions in the background. Similarly, although microeconomic decisions underlie macroeconomic phenomena, macroeconomic models do not necessarily focus on the optimizing behavior of households and firms; again, they sometimes leave that behavior in the background.

FYI

The Early Lives of Macroeconomists

How do people choose to become macroeconomists? There is no single path into the career. Here are the stories from some economists who later won Nobel Prizes for their work.[1]

Milton Friedman (Nobel 1976): "I graduated from college in 1932, when the United States was at the bottom of the deepest depression in its history before or since. The dominant problem of the time was economics. How to get out of the depression? How to reduce unemployment? What explained the paradox of great need on the one hand and unused resources on the other? Under the circumstances, becoming an economist seemed more relevant to the burning issues of the day than becoming an applied mathematician or an actuary."

James Tobin (Nobel 1981): "I was attracted to the field for two reasons. One was that economic theory is a fascinating intellectual challenge, on the order of mathematics or chess. I liked analytics and logical argument. . . . The other reason was the obvious relevance of economics to understanding and perhaps overcoming the Great Depression."

Franco Modigliani (Nobel 1985): "For a while it was thought that I should study medicine because my father was a physician. . . . I went to the registration window to sign up for medicine, but then I closed my eyes and thought of blood! I got pale just thinking about blood and decided under those conditions I had better keep away from medicine. . . . Casting about for something to do, I happened to get into some economics activities. I knew some German and was asked to translate from German into Italian some articles for one of the trade associations. Thus I began to be exposed to the economic problems that were in the German literature."

Robert Solow (Nobel 1987): "I came back [to college after being in the army] and, almost without thinking about it, signed up to finish my undergraduate degree as an economics major. The time was such that I had to make a decision in a hurry. No doubt I acted as if I were maximizing an infinite discounted sum of one-period utilities, but you couldn't prove it by me. To me it felt as if I were saying to myself: 'What the hell.' "

Robert Lucas (Nobel 1995): "In public school science was an unending and not very well organized list of things other people had discovered long ago. In college, I learned something about the process of scientific discovery, but what little I learned did not attract me as a career possibility. . . . What I liked thinking about were politics and social issues."

George Akerlof (Nobel 2001): "When I went to Yale, I was convinced that I wanted to be either an economist or an historian. Really, for me it was a distinction without a difference. If I was going to be an historian, then I would be an economic historian. And if I was to be an economist, I would consider history as the basis for my economics."

Edward Prescott (Nobel 2004): "Through discussion with [my father], I learned a lot about the way businesses operated. This was one reason why I liked my microeconomics course so much in my first year at Swarthmore College. The price theory that I learned in that course rationalized what I had learned from him about the way businesses operate. The other reason was the textbook used in that course, Paul A. Samuelson's *Principles of Economics*. I loved the way Samuelson laid out the theory in his textbook, so simply and clearly."

Edmund Phelps (Nobel 2006): "Like most Americans entering college, I started at Amherst College without a predetermined course of study and without even a career goal. My tacit assumption was that I would drift into the world of business — of money, doing something terribly smart. In the first year, though, I was awestruck by Plato, Hume, and James. I would probably have gone on to major in philosophy were it not that my father cajoled and pleaded with me to try a course in economics, which I did the second year. . . . I was hugely impressed to see that it was possible to subject the events in those newspapers I had read about to a formal sort of analysis."

Christopher Sims (Nobel 2011): "[My Uncle] Mark prodded me regularly, from about age 13 onward, to study economics. He gave me von Neumann and Morgenstern's *Theory of Games* for Christmas when I was in high school. When I took my first course in economics, I remember arguing with him over whether it was possible for the inflation rate to explode upward if the money supply were held constant. I took the monetarist position. He questioned whether I had a sound argument to support it. For years I thought he was having the opposite of his intended effect, and I studied no economics until my junior year of college. But as I began to doubt that I wanted to be immersed for my whole career in the abstractions of pure mathematics, Mark's efforts had left me with a pretty clear idea of an alternative."

[1] The first five quotations are from William Breit and Barry T. Hirsch, eds., *Lives of the Laureates*, 4th ed. (Cambridge, MA: MIT Press, 2004). The sixth, seventh, and ninth are from the Nobel website. The eighth is from Arnold Heertje, ed., *The Makers of Modern Economics,* vol. II (Aldershot, U.K.: Edward Elgar Publishing, 1995).

1-3 How This Book Proceeds

This book has five parts. This chapter and the next make up Part One, the “Introduction.” Chapter 2 discusses how economists measure economic variables, such as aggregate income, the inflation rate, and the unemployment rate.

Part Two, “Classical Theory: The Economy in the Long Run,” presents the classical model of how the economy works. The key assumption of the classical model is that prices are flexible. That is, with rare exceptions, the classical model assumes that markets clear. The assumption of price flexibility greatly simplifies the analysis, which is why we start with it. Yet because this assumption accurately describes the economy only in the long run, classical theory is best suited for analyzing a time horizon of at least several years.

Part Three, “Growth Theory: The Economy in the Very Long Run,” builds on the classical model. It maintains the assumptions of price flexibility and market clearing but adds a new emphasis on growth in the capital stock, the labor force, and technological knowledge. Growth theory is designed to explain how the economy evolves over a period of several decades.

Part Four, “Business Cycle Theory: The Economy in the Short Run,” examines the behavior of the economy when prices are sticky. The non-market-clearing model developed here is designed to analyze short-run issues, such as the reasons for economic fluctuations and the influence of government policy on those fluctuations. It is best suited for analyzing the changes in the economy we observe from month to month or from year to year.

Part Five, “Topics in Macroeconomic Theory and Policy,” covers material to supplement, reinforce, and refine our long-run and short-run analyses. Some chapters present advanced material of a somewhat theoretical nature, including macroeconomic dynamics, models of consumer behavior, and theories of firms’ investment decisions. Other chapters consider what role the government should have in the economy and discuss the debates over stabilization policy, government debt, and financial crises.

CHAPTER 2

The Data of Macroeconomics

Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

> It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.
>
> —Sherlock Holmes

Scientists, economists, and detectives have much in common: they all want to figure out what's going on in the world around them. To do this, they rely on theory and observation. They build theories to try to make sense of what they see happening. They then turn to more systematic observation to judge the theories' validity. Only when theory and evidence come into line do they feel they understand the situation. This chapter discusses the types of observation that economists use to develop and test their theories.

Casual observation is one source of information about what's happening in the economy. When you go shopping, you notice whether prices are rising, falling, or staying the same. When you look for a job, you learn whether firms are hiring. Every day, as we go about our lives, we participate in some aspect of the economy and get some sense of economic conditions.

A century ago, economists monitoring the economy had little more to go on than such casual observations. This fragmentary information made economic policymaking difficult. One person's anecdote would suggest the economy was moving in one direction, while another's would suggest otherwise. Economists needed some way to combine many individual experiences into a coherent whole. There was an obvious solution: as the old quip goes, the plural of "anecdote" is "data."

Today, economic data offer a systematic and objective source of information, and almost every day you can hear or read a story about some newly released statistic. Most of these statistics are produced by the government. Various government agencies survey households and firms to learn about their economic activity — how much they are earning, what they are buying, whether they have a job or are looking for work, what prices they are charging, how much they are producing, and so on. From these surveys, the agencies compute various statistics that summarize the state of the economy. Economists use these statistics to study the economy; policymakers use them to monitor developments and formulate policies.

This chapter focuses on the three statistics that economists and policymakers use most often. Gross domestic product, or GDP, tells us the nation’s total income and the total expenditure on its output of goods and services. The consumer price index, or CPI, measures the level of prices. The unemployment rate tells us the fraction of workers who are unemployed. In the following pages, we see how these statistics are computed and what they tell us about the economy.

2-1 Measuring the Value of Economic Activity: Gross Domestic Product

Gross domestic product, or **GDP**, is often considered the best measure of how well an economy is performing. In the United States, this statistic is computed every three months by the Bureau of Economic Analysis, a part of the U.S. Department of Commerce, from a large number of primary data sources. These primary sources include both (1) administrative data, which are byproducts of government functions such as tax collection, education programs, defense, and regulation and (2) statistical data, which come from government surveys of, for example, retail establishments, manufacturing firms, and farms. The purpose of GDP is to summarize all these data with a single number representing the dollar value of economic activity in a given period of time.

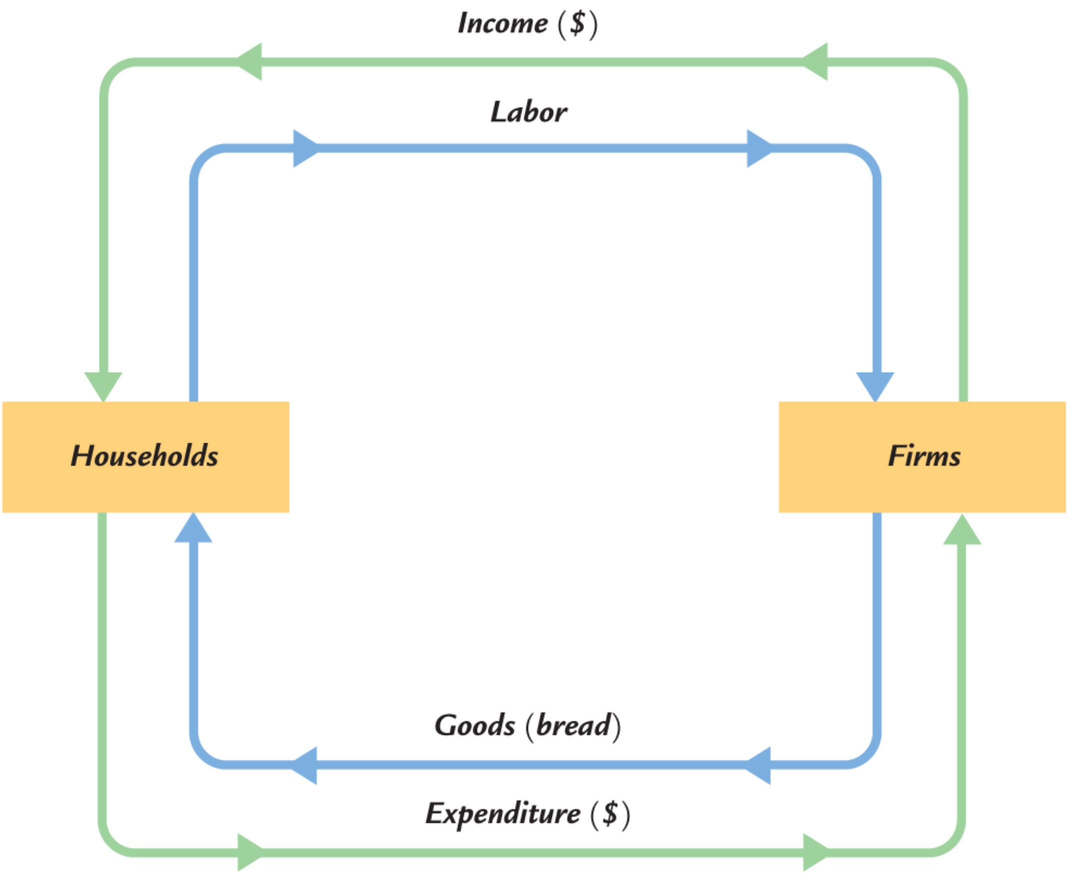
There are two ways to view this statistic. One way to view GDP is as *the total income of everyone in the economy*; another way is as *the total expenditure on the economy's output of goods and services*. From either viewpoint, it is clear why GDP is a gauge of economic performance. GDP measures something people care about — their incomes. Similarly, an economy with a large output of goods and services can better satisfy the demands of households, firms, and the government.

How can GDP measure both the economy's income and its expenditure on output? It can do so because these two quantities are really the same: for the economy as a whole, income must equal expenditure. That fact, in turn, follows from an even more fundamental one: because every transaction has a buyer and a seller, every dollar of expenditure by a buyer must become a dollar of income to a seller. When Jack paints Jill's house for $10,000, that $10,000 is income to Jack and expenditure by Jill. The transaction contributes $10,000 to GDP, regardless of whether we are adding up all income or all expenditure.

To understand the meaning of GDP more fully, we turn to **national income accounting**, the system used to measure GDP and many related statistics.

Income, Expenditure, and the Circular Flow

Imagine an economy that produces a single good, bread, from a single input, labor. Figure 2-1 illustrates all the economic transactions that occur between households and firms in this economy.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 2-1 The Circular Flow This figure illustrates the flows between firms and households in an economy that produces one good, bread, from one input, labor. The inner loop represents the flows of labor and bread: households sell their labor to firms, and the firms sell the bread they produce to households. The outer loop represents the corresponding flows of dollars: households pay the firms for the bread, and the firms pay wages and profit to the households. In this economy, GDP is both the total expenditure on bread and the total income from the production of bread.

The inner loop in Figure 2-1 represents the flows of bread and labor. The households sell their labor to the firms. The firms use the labor of their workers to produce bread, which the firms in turn sell to the households. Hence, labor flows from households to firms, and bread flows from firms to households.

The outer loop in Figure 2-1 represents the corresponding flow of dollars. The households buy bread from the firms. The firms use some of the revenue from these sales to pay the wages of their workers, and the remainder is the profit belonging to the owners of the firms (who themselves are part of the household sector). Hence, expenditure on bread flows from households to firms, and income in the form of wages and profit flows from firms to households.

GDP measures the flow of dollars in this economy. We can compute it in two ways. GDP is the total income from the production of bread, which equals the sum of wages and profit — the top half of the circular flow of dollars. GDP is also the total expenditure on purchases of bread — the bottom half of the circular flow of dollars. To compute GDP, we can look at either the flow of dollars from firms to households or the flow of dollars from households to firms.

These two ways of computing GDP must be equal because, by the rules of accounting, the expenditure of buyers on products is income to the sellers of those products. Every transaction that affects expenditure must affect income, and every transaction that affects income must affect expenditure. For example, suppose that a firm produces and sells one more loaf of bread to a household. Clearly this transaction raises total expenditure on bread, but it also has an equal effect on total income. If the firm produces the extra loaf without hiring any more labor (such as by making the production process more efficient), then profit increases. If the firm produces the extra loaf by hiring more labor, then wages increase. In both cases, expenditure and income increase equally.

FYI

Stocks and Flows

Many economic variables measure a quantity of something — a quantity of money, a quantity of goods, and so on. Economists distinguish between two types of quantity variables: stocks and flows. A **stock** is a quantity measured at a given point in time, whereas a **flow** is a quantity measured per unit of time.

A bathtub, shown in Figure 2-2, is the classic example used to illustrate stocks and flows. The amount of water in the tub is a stock: it is the quantity of water in the tub at a given point in time. The amount of water coming out of the faucet is a flow: it is the quantity of water being added to the tub per unit of time. Note that we measure stocks and flows in different units. We say that the bathtub contains 50 *gallons* of water but that water is coming out of the faucet at 5 *gallons per minute*.

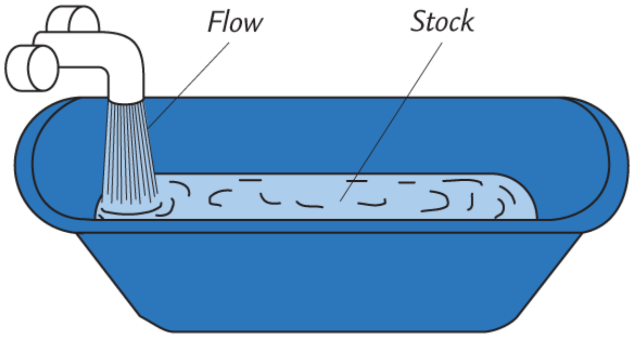


FIGURE 2-2 Stocks and Flows The amount of water in a bathtub is a stock: it is a quantity measured at a given moment in time. The amount of water coming out of the faucet is a flow: it is a quantity measured per unit of time.

GDP is probably the most important flow variable in economics: it tells us how many dollars are flowing around the economy's circular flow per unit of time. When someone says that the U.S. GDP is $20 trillion, this means that it is $20 trillion *per year*. (Equivalently, we could say that U.S. GDP is $634,000 per second.)

Stocks and flows are often related. In the bathtub example, these relationships are clear. The stock of water in the tub represents the accumulation of the flow out of the faucet, and the flow of water represents the change in the stock. When building theories to explain economic variables, it is often useful to determine whether the variables are stocks or flows and whether any relationships link them.

Here are some examples of related stocks and flows that we study in future chapters:

- A person's wealth is a stock; his income and expenditure are flows.
- The number of unemployed people is a stock; the number of people losing their jobs is a flow.

- The amount of capital in the economy is a stock; the amount of investment is a flow.
- The government debt is a stock; the government budget deficit is a flow.

Rules for Computing GDP

In an economy that produces only bread, we can compute GDP by adding up the total expenditure on bread. Real economies, however, include the production and sale of a vast number of goods and services. To compute GDP for such a complex economy, it is helpful to have a more precise definition: *gross domestic product (GDP) is the market value of all final goods and services produced within an economy in a given period of time*. To see how this definition is applied, let's discuss some of the rules that economists follow in constructing this statistic.

Adding Apples and Oranges

The U.S. economy produces many different goods and services — hamburgers, haircuts, cars, computers, and so on. GDP combines the value of these goods and services into a single measure. The diversity of products in the economy complicates the calculation of GDP because different products have different values.

Suppose, for example, that the economy produces four apples and three oranges. How do we compute GDP? We could simply add apples and oranges and conclude that GDP equals seven pieces of fruit. But this makes sense only if we think apples and oranges have equal value, which is generally not true. (This would be even clearer if the economy produces four watermelons and three grapes.)

To compute the total value of different goods and services, the national income accounts use market prices because these prices reflect how much people are willing to pay for a good or service. Thus, if apples cost \$0.50 each and oranges cost \$1.00 each, GDP would be

GDP=(Price of Apples×Quantity of Apples)+ (Price of Oranges×Quantity of Oranges)=($0.50×4)+($1.00×3)=$5.00.

$$\begin{aligned}\text{GDP} &= (\text{Price of Apples} \times \text{Quantity of Apples}) \\ &\quad + (\text{Price of Oranges} \times \text{Quantity of Oranges}) \\ &= (\$0.50 \times 4) + (\$1.00 \times 3) \\ &= \$5.00.\end{aligned}$$

GDP equals \$5.00 — the value of all the apples, \$2.00, plus the value of all the oranges, \$3.00.

Used Goods

When the Topps Company makes a pack of baseball cards and sells it for $2, that $2 is added to the nation's GDP. But when a collector sells a rare Mickey Mantle card to another collector for $500, that $500 is not part of GDP. GDP measures the value of *currently* produced goods and services. The sale of the Mickey Mantle card reflects the transfer of an asset, not an addition to the economy's income. Thus, the sale of used goods is not included in GDP.

Inventories

Imagine that a bakery hires workers to produce more bread, pays their wages, and then fails to sell the additional bread. How does this transaction affect GDP?

The answer depends on what happens to the unsold bread. Let's first suppose that the bread spoils. In this case, the firm has paid more in wages but has not received any additional revenue, so the firm's profit is reduced by the amount that wages have increased. Total expenditure in the economy hasn't changed because no one buys the bread. Total income hasn't changed either — although more is distributed as wages and less as profit. Because the transaction affects neither expenditure nor income, it does not alter GDP.

Now suppose, instead, that the bread is put into inventory (perhaps as frozen dough) to be sold later. In this case, the national income accounts treat the transaction differently. The owners of the firm are assumed to have "purchased" the bread for the firm's inventory, and the firm's profit is not reduced by the additional wages it has paid. Because the higher wages paid to the firm's workers raise total income, and the greater spending by the firm's owners on inventory raises total expenditure, the economy's GDP rises.

What happens later, when the firm sells the bread out of inventory? This case is similar to the sale of a used good. There is spending by bread consumers, but there is inventory disinvestment by the firm. This negative spending by the firm offsets the positive spending by consumers, so the sale out of inventory does not affect GDP.

The general rule is that when a firm increases its inventory of goods, this investment in inventory is counted as expenditure by the firm owners. Thus, production for inventory increases GDP just as much as does production for final sale. A sale out of inventory, however, combines positive spending (the purchase) and negative spending (inventory disinvestment), so it does not affect GDP. This treatment of inventories ensures that GDP reflects the economy's current production of goods and services.

Intermediate Goods

Many goods are produced in stages: raw materials are processed into intermediate goods by one firm and then sold to another firm for final processing. How should we treat such products when computing GDP? For example, suppose a cattle rancher sells one-quarter pound of meat to McDonald's for $1, and then McDonald's sells you a hamburger for $3. Should GDP include both the meat and the hamburger (a total of $4) or just the hamburger ($3)?

The answer is that GDP includes only the value of final goods. Thus, the hamburger is included in GDP, but the meat is not: GDP increases by $3, not by $4. The reason is that the value of intermediate goods is already included as part of the market price of the final goods in which they are used. To add the intermediate goods to the final goods would be double counting — that is, the meat would be counted twice. Hence, GDP is the total value of final goods and services produced.

One way to compute the value of all final goods and services is to sum the value added at each stage of production. The **value added** of a firm equals the value of the firm's output less the value of the intermediate goods that the firm purchases. In the case of the hamburger, the value added of the rancher is $1 (assuming that the rancher bought no intermediate goods), and the value added of McDonald's is $3−$1, or $2. $3 − $1, or $2. Total value added is $1+$2, $1 + $2, which equals $3. For the economy as a whole, the sum of all value added must equal the value of all final goods and services. Hence, GDP is also the total value added of all firms in the economy.

Imputations

Although most goods and services are valued at their market prices when computing GDP, some are not sold in the marketplace and therefore do not have market prices. If GDP is to include these goods and services, we must use an estimate of their value. Such an estimate is called an **imputed value**.

Imputations are especially important for determining the value of housing. A person who rents a house is buying housing services and providing income for the landlord; the rent is part of GDP, both as expenditure by the renter and as income for the landlord. Some people, however, own their homes. They do not pay rent to a landlord, but they are enjoying housing services similar to those that renters purchase. To take account of the housing services enjoyed by homeowners, GDP includes the "rent" that these homeowners "pay" to themselves. Of course, homeowners do not in fact pay themselves this rent. The Department of Commerce estimates what the market rent for a house would be if it were rented and includes that imputed rent as part of GDP. This imputed rent is included both in the homeowner's expenditure and in the homeowner's income.

Imputations also arise in valuing government services. For example, police officers, firefighters, and senators provide services to the public. Assigning a value to these services is difficult because they are not sold in a marketplace and therefore do not have a market price. The national income accounts include these services in GDP by valuing them at their cost. That is, the wages of these public servants are used to measure the value of their output.

In many cases, an imputation is called for in principle but, to keep things simple, is not made in practice. Because GDP includes the imputed rent on owner-occupied houses, one might expect it also to include the imputed rent on cars, lawn mowers, jewelry, and other durable goods owned by households. Yet the value of these rental services is left out of GDP. In addition, some of the output of the economy is produced and consumed at home and never enters the marketplace. For example, meals cooked at home are similar to meals cooked at a restaurant, yet the value added when a person prepares a meal at home is left out of GDP.

Finally, no imputation is made for the value of goods and services sold in the *underground economy*. The underground economy is the part of the economy that people hide from the government either because they wish to evade taxation or because the activity is illegal. Examples include domestic workers paid “off the books” and the illegal drug trade. The size of the underground economy varies widely from country to country. In the United States, the underground economy is estimated to be less than 10 percent of the official economy, whereas in some developing nations, such as Thailand, Nigeria, and Bolivia, the underground economy is more than half as large as the official one.

Because the imputations necessary for computing GDP are only approximate, and because the value of many goods and services is left out altogether, GDP is an imperfect measure of economic activity. These imperfections are most problematic when comparing standards of living across countries. Yet as long as the magnitude of these imperfections remains fairly constant over time, GDP is useful for comparing economic activity from year to year.

Real GDP Versus Nominal GDP

Economists use the rules just described to compute GDP, which values the economy’s total output of goods and services. But is GDP a good measure of economic well-being? Consider once again the economy that produces only apples and oranges. In this economy, GDP is the sum of the value of all the apples produced and the value of all the oranges produced. That is,

GDP=(Price of Apples×Quantity of Apples)+ (Price of Oranges×Quantity of Oranges).

$$\begin{aligned}\text{GDP} &= (\text{Price of Apples} \times \text{Quantity of Apples}) \\ &\quad + (\text{Price of Oranges} \times \text{Quantity of Oranges}).\end{aligned}$$

Economists call the value of goods and services measured at current prices **nominal GDP**. Notice that nominal GDP can increase either because prices rise or because quantities rise.

It is easy to see that GDP computed this way is not a good gauge of economic well-being. That is, this measure does not accurately reflect how well the economy can satisfy the demands of households, firms, and the government. If all prices doubled without any change in quantities, nominal GDP would double. Yet it would be misleading to say that the economy’s ability to satisfy demands has doubled because the quantity of every good produced remains the same.

A better measure of economic well-being would tally the economy’s output of goods and services without being influenced by changes in prices. For this purpose, economists use **real GDP**, which is the value of goods and services measured using a constant set of prices. That is, real GDP shows what would have happened to expenditure on output if quantities had changed but prices had not.

To see how real GDP is computed, imagine we want to compare output in 2017 with output in subsequent years for our apple-and-orange economy. We could begin by choosing a set of prices, called *base-year prices*, such as the prices that prevailed in 2017. Goods and services are then added up using these base-year prices to value the different goods in each year. Real GDP for 2017 would be

Real GDP=(2017 Price of Apples×2017 Quantity of Apples)+ (2017 Price of Oranges×2017 Quantity of Oranges).

$$\begin{aligned}\text{Real GDP} &= (\text{2017 Price of Apples} \times \text{2017 Quantity of Apples}) \\ &\quad + (\text{2017 Price of Oranges} \times \text{2017 Quantity of Oranges}).\end{aligned}$$

Similarly, real GDP in 2018 would be

Real GDP=(2017 Price of Apples×2018 Quantity of Apples)+ (2017 Price of Oranges×2018 Quantity of Oranges).

$$\begin{aligned}\text{Real GDP} &= (\text{2017 Price of Apples} \times \text{2018 Quantity of Apples}) \\ &\quad + (\text{2017 Price of Oranges} \times \text{2018 Quantity of Oranges}).\end{aligned}$$

And real GDP in 2019 would be

Real GDP=(2017 Price of Apples×2019 Quantity of Apples)+ (2017 Price of Oranges×2019 Quantity of Oranges).

$$\begin{aligned}\text{Real GDP} &= (\text{2017 Price of Apples} \times \text{2019 Quantity of Apples}) \\ &\quad + (\text{2017 Price of Oranges} \times \text{2019 Quantity of Oranges}).\end{aligned}$$

Notice that 2017 prices are used to compute real GDP for all three years. With the prices held constant, real GDP varies from year to year only if the quantities produced vary. Because a society’s ability to provide economic satisfaction for its members ultimately depends on the quantities of goods and services produced, real GDP provides a better measure of economic well-being than does nominal GDP.

The GDP Deflator

From nominal GDP and real GDP, we can compute a third statistic: the GDP deflator. The **GDP deflator**, also called the *implicit price deflator for GDP*, is the ratio of nominal GDP to real GDP:

$$\text{GDP Deflator} = \frac{\text{Nominal GDP}}{\text{Real GDP}}.$$

The GDP deflator reflects what's happening to the overall level of prices in the economy.

To better understand this, consider again an economy with only one good, bread. If *P* is the price of bread and *Q* is the quantity sold, then nominal GDP is the total number of dollars spent on bread in that year, $P \times Q$. Real GDP is the number of loaves of bread produced in that year times the price of bread in some base year, $P_{\text{base}} \times Q$. The GDP deflator is the price of bread in that year relative to the price of bread in the base year, P/P_{base}.

The definition of the GDP deflator allows us to separate nominal GDP into two parts: one part measures quantities (real GDP) and the other measures prices (the GDP deflator). That is,

$$\text{Nominal GDP} = \text{Real GDP} \times \text{GDP Deflator}.$$

Nominal GDP measures the current dollar value of the output of the economy. Real GDP measures output valued at constant prices. The GDP deflator measures the price of output relative to its price in the base year. We can also write this equation as

$$\text{Real GDP} = \frac{\text{Nominal GDP}}{\text{GDP Deflator}}.$$

In this form, you can see how the deflator earns its name: it is used to deflate (that is, take inflation out of) nominal GDP to yield real GDP.

Chain-Weighted Measures of Real GDP

We have been discussing real GDP as if the prices used to compute this measure never change from their base-year values. If this were truly the case, over time the prices would become more and more dated. For instance, the price of computers has fallen substantially in recent years, while the price of a year at college has risen. When valuing the production of computers and education, it would be misleading to use the prices that prevailed ten or twenty years ago.

To solve this problem, the Bureau of Economic Analysis used to periodically update the prices used to compute real GDP. About every five years, a new base year was chosen. The prices were then held fixed and used to measure year-to-year changes in the production of goods and services until the base year was updated once again.

In 1995, the Bureau announced a new policy for dealing with changes in the base year. In particular, it now uses *chain-weighted* measures of real GDP. With these new measures, the base year changes continuously over time. In essence, average prices in 2017 and 2018 are used to measure real growth from 2017 to 2018, average prices in 2018 and 2019 are used to measure real growth from 2018 to 2019, and so on. These various year-to-year growth rates are then put together to form a “chain” that can be used to compare the output of goods and services between any two dates.

This new chain-weighted measure of real GDP is better than the more traditional measure because it ensures that the prices used to compute real GDP are never far out of date. For most purposes, however, the differences are not significant. It turns out that the two measures of real GDP are highly correlated with each other. As a practical matter, both measures of real GDP reflect the same thing: economy-wide changes in the production of goods and services.

FYI

Two Helpful Hints for Working with Percentage Changes

For manipulating many relationships in economics, there is an arithmetic fact that is useful to learn: *The percentage change of a product of two variables is approximately the sum of the percentage changes in each of the variables*.

Consider an example. Let *P* denote the GDP deflator and *Y* denote real GDP. Nominal GDP is P×Y. $P \times Y$. Applying the arithmetic fact, we get

Percentage Change in (P×Y)≈(Percentage Change in P)+(Percentage Change in Y).

$$\text{Percentage Change in } (P \times Y) \approx (\text{Percentage Change in } P) + (\text{Percentage Change in } Y).$$

For instance, suppose that in one year, real GDP is 100 and the GDP deflator is 2; the next year, real GDP is 103 and the GDP deflator is 2.1. We can calculate that real GDP rose by 3 percent and that the GDP deflator rose by 5 percent. Nominal GDP rose from 200 the first year to 216.3 the second year, an increase of 8.15 percent. Notice that the growth in nominal GDP (8.15 percent) is approximately the sum of the growth in the GDP deflator (5 percent) and the growth in real GDP (3 percent).[1]

A second arithmetic fact follows as a corollary to the first: *The percentage change of a ratio is approximately the percentage change in the numerator minus the percentage change in the denominator.* Again, consider an example. Let *Y* denote GDP and *L* denote the population, so that *Y*/*L* is GDP per person. The second fact states that

Percentage Change in (Y/L)≈(Percentage Change in Y)−(Percentage Change in L).

$$\text{Percentage Change in } (Y/L) \approx (\text{Percentage Change in } Y) - (\text{Percentage Change in } L).$$

For instance, suppose that in the first year, *Y* is 100,000 and *L* is 100, so *Y*/*L* is 1,000; in the second year, *Y* is 110,000 and *L* is 103, so *Y*/*L* is 1,068. Notice that the growth in GDP per person (6.8 percent) is approximately the growth in income (10 percent) minus the growth in population (3 percent).

The Components of Expenditure

Economists and policymakers care not only about the economy's total output of goods and services but also about the allocation of this output among alternative uses. The national income accounts divide GDP into four broad categories of spending:

- Consumption (*C*)
- Investment (*I*)
- Government purchases (*G*)
- Net exports (*NX*).

Thus, letting *Y* stand for GDP,

Y=C+I+G+NX.$$Y = C + I + G + NX.$$

GDP is the sum of consumption, investment, government purchases, and net exports. Each dollar of GDP falls into one of these categories. This equation is an *identity* — an equation that must hold because of the way the variables are defined. It is called the **national income accounts identity**.

Consumption consists of household expenditures on goods and services. Goods are tangible items, and they in turn are split into durables and nondurables. Durable goods are goods that last a long time, such as cars and TVs. Nondurable goods are goods that last only a short time, such as food and clothing. Services include various intangible items that consumers buy, such as haircuts and doctor visits.

Investment consists of items bought for future use. Investment is divided into three subcategories: business fixed investment, residential fixed investment, and inventory investment. Business fixed investment, also called nonresidential fixed investment, is the purchase by firms of new structures, equipment, and intellectual property products. (Intellectual property products include software, research and development, and entertainment, literary, and artistic originals.) Residential investment is the purchase of new housing by households and landlords. Inventory investment is the increase in firms' inventories of goods. (If inventories are falling, inventory investment is negative.)

Government purchases are the goods and services bought by federal, state, and local governments. This category includes such items as military equipment, highways, and the services provided by government workers. It does not include transfer payments to individuals, such as Social Security and welfare. Because transfer payments reallocate existing income and are not made in exchange for goods and services, they are not part of GDP.

The last category, **net exports**, accounts for trade with other countries. Net exports are the value of goods and services sold to other countries (exports) minus the value of goods and services that other countries sell to us (imports). Net exports are positive when the value of our exports is greater than the value of our imports and negative when the value of our imports is greater than the value of our exports. Net exports represent the net expenditure from abroad on our goods and services, which provides income for domestic producers.

FYI

What Is Investment?

Newcomers to macroeconomics are sometimes confused by how macroeconomists use familiar words in new and specific ways. One example is the term *investment*. The confusion arises because what looks like investment for an individual may not be investment for the economy as a whole. The general rule is that the economy's investment does not include purchases that merely reallocate existing assets among different individuals. Investment, as macroeconomists use the term, creates a new physical asset, called *capital*, which can be used in future production.

Let's consider some examples. Suppose we observe these two events:

- Smith buys himself a 100-year-old Victorian house.
- Jones builds herself a brand-new contemporary house.

What is total investment here? Two houses, one house, or zero?

A macroeconomist seeing these two transactions counts only the Jones house as investment. Smith's transaction has not created new housing for the economy; it has merely reallocated existing housing to Smith from the previous owner. By contrast, because Jones has added new housing to the economy, her new house is counted as investment.

Similarly, consider these two events:

- Gates buys $5 million in IBM stock from Buffett on the New York Stock Exchange.
- General Motors sells $10 million in stock to the public and uses the proceeds to build a new car factory.

Here, investment is $10 million. The first transaction reallocates ownership of shares in IBM from Buffett to Gates; the economy's stock of capital is unchanged, so there is no investment as macroeconomists use the term. By contrast, because General Motors is using some of the economy's output of goods and services to add to its stock of capital, its new factory is counted as investment.

CASE STUDY

GDP and Its Components

In 2016, the GDP of the United States totaled about $18.6 trillion. This number is so large that it is hard to comprehend. We can make it easier to understand by dividing it by the 2016 U.S. population of 323 million. In this way, we obtain GDP per person — the amount of expenditure for the average American — which equaled $57,638.

How did this GDP get used? Table 2-1 shows that about two-thirds of it, or $39,677 per person, was spent on consumption. Investment was $9,461 per person. Government purchases were $10,113 per person, $2,256 of which was spent by the federal government on national defense.

TABLE 2-1 GDP and the Components of Expenditure: 2016

	Total (billions of dollars)	Per Person (dollars)
Gross Domestic Product	18,624	57,638
Consumption	12,821	39,677
Nondurable goods	2,710	8,388
Durable goods	1,411	4,367
Services	8,699	26,922
Investment	3,057	9,461
Nonresidential fixed investment	2,316	7,168
Residential fixed investment	706	2,185
Inventory investment	35	109
Government Purchases	3,268	10,113
Federal	1,231	3,811

Defense	729	2,256
Nondefense	503	1,555
State and local	2,036	6,302
Net Exports	−521	−1,613
Exports	2,215	6,854
Imports	2,736	8,467
Data from: U.S. Department of Commerce, U.S. Census Bureau.		

The average American bought $8,467 of goods imported from abroad and produced $6,854 of goods that were exported to other countries. Because the average American imported more than he exported, net exports were negative. Furthermore, because the average American earned less from selling to foreigners than he spent on foreign goods, he must have financed the difference by taking out loans from foreigners (or, equivalently, by selling them some of his assets). Thus, the average American borrowed $1,613 from abroad in 2016. ■

Other Measures of Income

The national income accounts include other measures of income that differ slightly in definition from GDP. It is important to be aware of the various measures because economists and the media often refer to them.

To see how the alternative measures of income relate to one another, we start with GDP and modify it in various ways. To obtain *gross national product* (*GNP*), we add to GDP receipts of factor income (wages, profit, and rent) from the rest of the world and subtract payments of factor income to the rest of the world:

GNP=GDP+Factor Payments from Abroad−Factor Payments to Abroad.

$$\text{GNP} = \text{GDP} + \text{Factor Payments from Abroad} - \text{Factor Payments to Abroad.}$$

Whereas GDP measures the total income produced *domestically*, GNP measures the total income earned by *nationals* (residents of a nation). For instance, if a Japanese resident owns an apartment building in New York, the rental income he earns is part of U.S. GDP because it is earned in the United States. But because this rental income is a factor payment to abroad, it is not part of U.S. GNP. In the United States, factor payments from abroad and factor payments to abroad are similar in size — each representing about 4 percent of GDP — so GDP and GNP are quite close.

To obtain *net national product* (*NNP*), we subtract from GNP the depreciation of capital — the amount of the economy's stock of plants, equipment, and residential structures that wears out during the year:

$$\text{NNP} = \text{GNP} - \text{Depreciation.}$$
$$\text{NNP} = \text{GNP} - \text{Depreciation.}$$

In the national income accounts, depreciation is called the *consumption of fixed capital*. It equals about 16 percent of GNP. Because the depreciation of capital is a cost of producing the output of the economy, subtracting depreciation shows the net result of economic activity.

Net national product is approximately equal to another measure called *national income*. The two differ by a small correction called the *statistical discrepancy*, which arises because different data sources may not be completely consistent:

$$\text{National Income} = \text{NNP} - \text{Statistical Discrepancy.}$$
$$\text{National Income} = \text{NNP} - \text{Statistical Discrepancy.}$$

National income measures how much everyone in the economy has earned.

The national income accounts divide national income into six components, depending on who earns the income. The six categories, and the percentage of national income paid in each category in 2016, are as follows:

- *Compensation of employees* (62 percent). The wages and fringe benefits earned by workers.
- *Proprietors' income* (8 percent). The income of noncorporate businesses, such as small farms, mom-and-pop stores, and law partnerships.
- *Rental income* (4 percent). The income that landlords receive, including the imputed rent that homeowners "pay" to themselves, less expenses, such as depreciation.
- *Corporate profits* (13 percent). The income of corporations after payments to their workers and creditors.
- *Net interest* (4 percent). The interest domestic businesses pay minus the interest they receive, plus interest earned from foreigners.
- *Taxes on production and imports* (9 percent). Certain taxes on businesses, such as sales taxes, less offsetting business subsidies. These taxes place a wedge between the price that consumers pay for a good and the price that firms receive.

A series of adjustments take us from national income to *personal income*, the amount of income that households and noncorporate businesses receive. Four of these adjustments are most important. First, we subtract taxes on production and imports because these taxes never enter anyone's income. Second, we reduce national income by the amount that corporations earn but do not pay out, either because the corporations are retaining earnings or because they are paying taxes to the government. This adjustment is made by subtracting corporate profits (which equal the sum of corporate taxes, dividends, and retained earnings) and adding back dividends. Third, we increase national income by the net amount the government pays out in transfer payments. This adjustment equals government transfers to individuals minus social insurance contributions paid to the government. Fourth, we adjust national income to include the interest that households earn rather than the interest that businesses pay. This adjustment is made by adding personal interest income and subtracting net interest. (The difference between personal interest and net interest arises in part because interest on the government debt is part of the interest that households earn but is not part of the interest that businesses pay out.) Thus,

Personal Income=National Income− Indirect Business Taxes− Corporate Profits− Social Insurance Contributions− Net Interest+ Dividends+ Government Transfers to Individuals+ Personal Interest Income.

$$\begin{aligned}\text{Personal Income} \quad = \quad & \text{National Income} \\ & - \text{Indirect Business Taxes} \\ & - \text{Corporate Profits} \\ & - \text{Social Insurance Contributions} \\ & - \text{Net Interest} \\ & + \text{Dividends} \\ & + \text{Government Transfers to Individuals} \\ & + \text{Personal Interest Income.}\end{aligned}$$

Next, if we subtract personal taxes, we obtain *disposable personal income*:

Disposable Personal Income=Personal Income−Personal Taxes.

$$\text{Disposable Personal Income} = \text{Personal Income} - \text{Personal Taxes.}$$

We are interested in disposable personal income because it is the amount households and noncorporate businesses have available to spend after satisfying their tax obligations to the government.

Seasonal Adjustment

Because real GDP and the other measures of income reflect how well the economy is performing, economists are interested in studying the quarter-to-quarter fluctuations in these variables. Yet when we start to do so, one fact leaps out: all these measures of income exhibit a regular seasonal pattern. The output of the economy rises during the year, reaching a peak in the fourth quarter (October, November, and December) and then falling in the first quarter (January, February, and March) of the next year. These regular seasonal changes are substantial. From the fourth quarter to the first quarter, real GDP falls on average about 8 percent.[2]

It is not surprising that real GDP follows a seasonal cycle. Some of these changes are attributable to changes in our ability to produce: for example, building homes is more difficult during the cold weather of winter than during other seasons. In addition, people have seasonal tastes: they have preferred times for activities like vacations and Christmas shopping.

When economists study fluctuations in real GDP and other economic variables, they often want to eliminate the portion of fluctuations due to predictable seasonal changes. You will find that most of the economic statistics reported are *seasonally adjusted*. This means that the data have been adjusted to remove the regular seasonal fluctuations. (The precise statistical procedures used are too elaborate to discuss here, but in essence they involve subtracting those changes in income that are predictable just from the change in season.) Therefore, when you observe a rise or fall in real GDP or any other data series, you must look beyond the seasonal cycle for the explanation.[3]

2-2 Measuring the Cost of Living: The Consumer Price Index

A dollar today doesn't buy as much as it did twenty years ago. The cost of almost everything has gone up. This increase in the overall level of prices is called *inflation*, and the percentage change in the price level from one period to the next is called the *inflation rate*. Inflation is a primary concern of economists and policymakers. In later chapters, we examine the causes and effects of inflation. Here we discuss how economists measure changes in the cost of living.

The Price of a Basket of Goods

The most commonly used measure of the level of prices is the **consumer price index (CPIS)**. The Bureau of Labor Statistics (BLS) has the job of computing the CPI. It begins by collecting the prices of thousands of goods and services. Just as GDP turns the quantities of many goods and services into a single number measuring the value of production, the CPI turns the prices of many goods and services into a single index measuring the overall level of prices.

How should economists aggregate the many prices in the economy into a single index that reliably measures the price level? They could simply compute an average of all prices. But this approach would treat all goods and services equally. Because people buy more chicken than caviar, the price of chicken should have a greater weight in the CPI than the price of caviar. The BLS weights different items by computing the price of a basket of goods and services purchased by a typical consumer. The CPI is the price of this basket of goods and services relative to the price of the same basket in some base year.

For example, suppose that the typical consumer buys five apples and two oranges every month. Then the basket of goods consists of five apples and two oranges, and the CPI is

CPI=(5×Current Price of Apples)+(2×Current Price of Oranges)(5×2017 Price of Apples)+(2×2017 Price of Oranges)

$$\text{CPI} = \frac{(5 \times \text{Current Price of Apples}) + (2 \times \text{Current Price of Oranges})}{(5 \times \text{2017 Price of Apples}) + (2 \times \text{2017 Price of Oranges})}$$

In this CPI, 2017 is the base year. The index tells us how much it costs now to buy five apples and two oranges relative to how much it cost to buy the same basket of fruit in 2017.

The consumer price index is the most closely watched index of prices, but it is not the only such index. Another is the *producer price index*, which measures the price of a typical basket of goods bought by firms rather than consumers. In addition to these overall price indexes, the BLS computes price indexes for specific types of goods, such as food, housing, and energy. Another statistic, sometimes called *core inflation*, measures the increase in price of a consumer basket that excludes food and energy products. Because food and energy prices exhibit substantial short-run volatility, core inflation is sometimes viewed as a better gauge of ongoing inflation trends.

How the CPI Compares to the GDP and PCE Deflators

Earlier in this chapter, we saw another measure of prices — the implicit price deflator for GDP, which is the ratio of nominal GDP to real GDP. The GDP deflator and the CPI give somewhat different information about what's happening to the overall level of prices in the economy. There are three key differences between the two measures.

The first difference is that the GDP deflator measures the prices of all goods and services produced, whereas the CPI measures the prices of only the goods and services bought by consumers. Thus, an increase in the price of goods bought only by firms or the government will show up in the GDP deflator but not in the CPI.

The second difference is that the GDP deflator includes only those goods produced domestically. Imported goods are not part of GDP and do not show up in the GDP deflator. Hence, an increase in the price of Toyotas made in Japan and sold in this country affects the CPI because the Toyotas are bought by consumers, but it does not affect the GDP deflator.

The third and most subtle difference results from the way the two measures aggregate the many prices in the economy. The CPI assigns fixed weights to the prices of different goods, whereas the GDP deflator assigns changing weights. In other words, the CPI is computed using a fixed basket of goods, whereas the GDP deflator allows the basket of goods to change over time as the composition of GDP changes. The following example shows how these approaches differ. Suppose that major frosts destroy the nation's orange crop. The quantity of oranges produced falls to zero, and the price of the few oranges that remain on grocers' shelves is driven sky high. Because oranges are no longer part of GDP, the increase in the price of oranges does not show up in the GDP deflator. But because the CPI is computed with a fixed basket of goods that includes oranges, the increase in the price of oranges causes a substantial rise in the CPI.

Economists call a price index with a fixed basket of goods a *Laspeyres index* and a price index with a changing basket a *Paasche index*. Economic theorists have studied the properties of these different types of price indexes to determine which is a better measure of the cost of living. The answer, it turns out, is that neither is clearly superior. When prices of different goods are changing by different amounts, a Laspeyres (fixed basket) index tends to overstate the increase in the cost of living because it does not take into account the fact that consumers have the opportunity to substitute less expensive goods for more expensive ones. By contrast, a Paasche (changing basket) index tends to understate the increase in the cost of living. Although it accounts for the substitution of alternative goods, it does not reflect the reduction in consumers' welfare that result from such substitutions.

The example of the destroyed orange crop shows the problems with Laspeyres and Paasche price indexes. Because the CPI is a Laspeyres index, it overstates the impact of the increase in orange prices on consumers: by using a fixed basket of goods, it ignores consumers' ability to substitute apples for oranges. By contrast, because the GDP deflator is a Paasche index, it understates the impact on consumers: the GDP deflator shows no rise in prices, yet surely the higher price of oranges makes consumers worse off.[4]

In addition to the CPI and the GDP deflator, another noteworthy measure of inflation is the implicit price deflator for personal consumption expenditures (PCE), or **PCE deflator**. The PCE deflator is calculated like the GDP deflator but, rather than being based on all of GDP, it is based on just the consumption component. That is, the PCE deflator is the ratio of nominal consumer spending to real consumer spending.

The PCE deflator resembles the CPI in some ways and the GDP deflator in others. Like the CPI, the PCE deflator includes only the prices of goods and services that consumers buy; it excludes the prices of goods and services that are part of investment and government purchases. Also like the CPI, the PCE deflator includes the prices of imported goods. But like the GDP deflator, the PCE deflator allows the basket of goods to change over time as the composition of consumer spending changes. Because of this mix of attributes, the Federal Reserve uses the PCE deflator as its preferred gauge of how quickly prices are rising.

Luckily, the differences among these various measures of inflation are usually small in practice. Figure 2-3 shows inflation as measured by the CPI, the GDP deflator, and the PCE deflator for each year from 1948 to 2016. All three measures usually tell the same story about how quickly prices are rising.

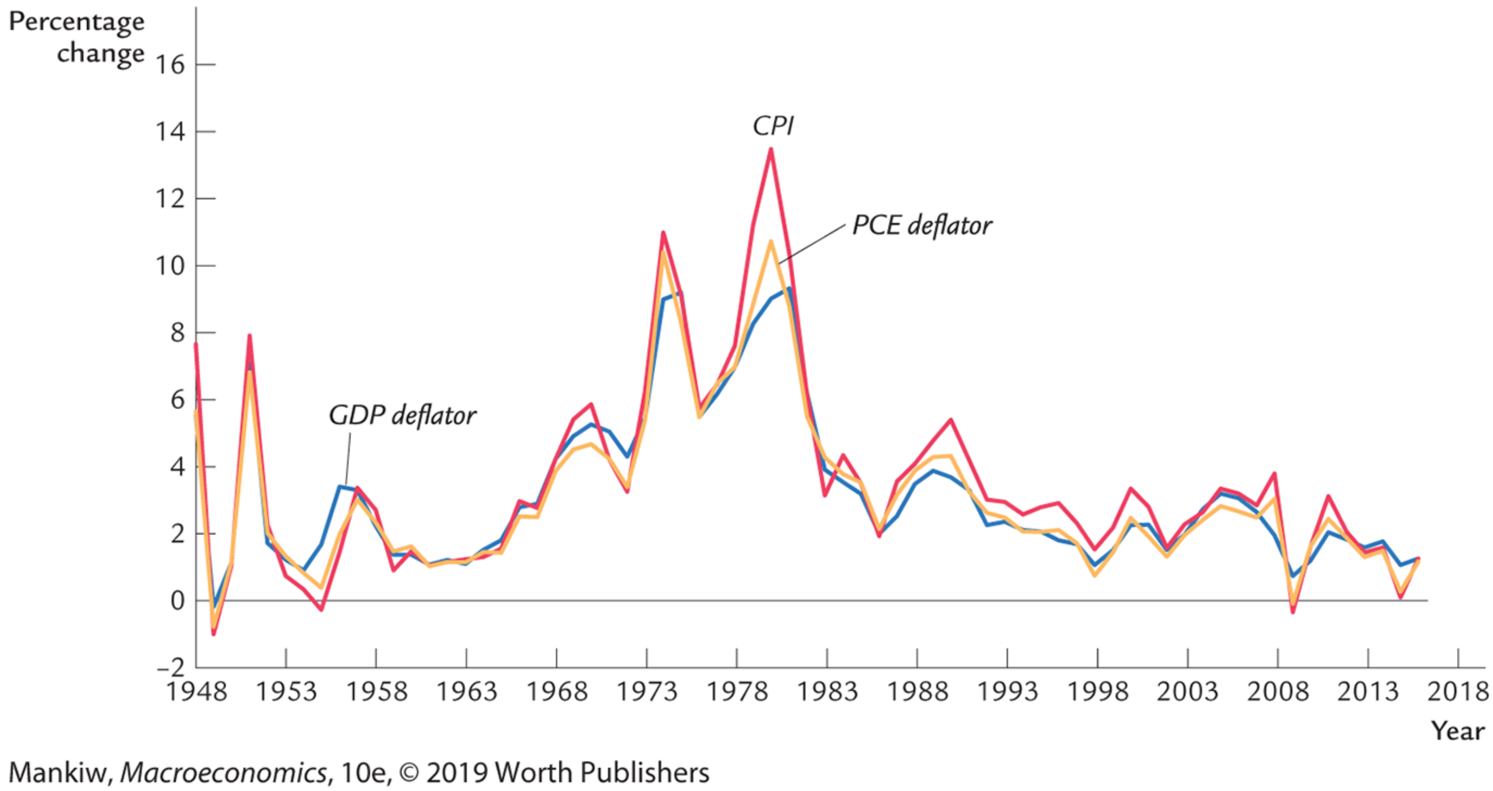


FIGURE 2-3 Three Measures of Inflation This figure shows the percentage change in the CPI, the GDP deflator, and the PCE deflator for every year from 1948 to 2016. These measures of prices diverge at times, but they usually tell the same story about how quickly prices are rising. Both the CPI and the GDP deflator show that prices rose slowly in most of the 1950s and 1960s, that they rose much more quickly in the 1970s, and that they have risen slowly again since the mid-1980s.

Data from: U.S. Department of Commerce, U.S. Department of Labor.

Does the CPI Overstate Inflation?

The consumer price index is a closely watched measure of inflation. Policymakers in the Federal Reserve monitor it, along with many other variables, when setting monetary policy. In addition, many laws and private contracts have cost-of-living allowances, called *COLAs*, which use the CPI to adjust for changes in the price level. For instance, Social Security benefits are adjusted automatically every year so that inflation will not erode the living standard of the elderly.

Because so much depends on the CPI, it is important to ensure that this measure of the price level is accurate. Many economists believe that, for a number of reasons, the CPI tends to overstate inflation.

One problem is the substitution bias we have already discussed. Because the CPI measures the price of a fixed basket of goods, it does not reflect the ability of consumers to substitute toward goods whose relative prices have fallen. Thus, when relative prices change, the true cost of living rises less rapidly than does the CPI.

A second problem is the introduction of new goods. When a new good is introduced into the marketplace, consumers are better off because they have more products from which to choose. In effect, the introduction of new goods increases the real value of the dollar. Yet this increase in the purchasing power of the dollar is not reflected in a lower CPI.

A third problem is unmeasured changes in quality. When a firm changes the quality of a good it sells, not all of the good’s price change reflects a change in the cost of living. The BLS does its best to account for changes in the quality of goods over time. For example, if Ford increases the horsepower of a particular car model from one year to the next, the CPI will reflect the change: the quality-adjusted price of the car will not rise as fast as the unadjusted price. Yet many changes in quality, such as comfort or safety, are hard to measure. If unmeasured quality improvement (rather than unmeasured quality deterioration) is typical, then the measured CPI rises faster than it should.

In 1995, the Senate Finance Committee appointed a panel of economists to study the magnitude of the measurement error in the CPI. The panel concluded that the CPI was biased upward by 0.8 to 1.6 percentage points per year, with their “best estimate” being 1.1 percentage points. This report led to some changes in the way the CPI is calculated, so the bias is now thought to be under 1 percent point. The CPI still overstates inflation, but not by as much as it once did.[5]

2-3 Measuring Joblessness: The Unemployment Rate

One aspect of economic performance is how well an economy uses its resources. Because an economy's workers are its chief resource, keeping workers employed is a paramount concern of economic policymakers. The unemployment rate is the statistic that measures the percentage of those people wanting to work who do not have jobs. Every month, the U.S. Bureau of Labor Statistics (BLS) computes the unemployment rate and many other statistics that economists and policymakers use to monitor developments in the labor market.

The Household Survey

The unemployment rate comes from a survey of about 60,000 households called the Current Population Survey. These households include about 110,000 individuals. Based on the responses to survey questions, each adult (age 16 and older) is placed into one of three categories:

- *Employed.* This category includes those who at the time of the survey worked as paid employees, worked in their own business, or worked as unpaid workers in a family member's business. It also includes those who were not working but who had jobs from which they were temporarily absent because of, for example, vacation, illness, or bad weather.
- *Unemployed.* This category includes those who were not employed, were available for work, and had tried to find employment during the previous four weeks. It also includes those waiting to be recalled to a job from which they had been laid off.
- *Not in the labor force.* This category includes those who fit neither of the first two categories, such as a full-time student, homemaker, or retiree.

Notice that a person who wants a job but has given up looking — a *discouraged worker* — is counted as not being in the labor force.

The **labor force** is the sum of the employed and unemployed, and the **unemployment rate** is the percentage of the labor force that is unemployed. That is,

Labor Force=Number of Employed+Number of Unemployed

$$\text{Labor Force} = \text{Number of Employed} + \text{Number of Unemployed}$$

and

Unemployment Rate=Number of UnemployedLabor Force×100.

$$\text{Unemployment Rate} = \frac{\text{Number of Unemployed}}{\text{Labor Force}} \times 100.$$

A related statistic is the **labor-force participation rate**, the percentage of the adult population that is in the labor force:

Labor-Force Participation Rate=Labor ForceAdult Population×100.

$$\text{Labor-Force Participation Rate} = \frac{\text{Labor Force}}{\text{Adult Population}} \times 100.$$

The BLS computes these statistics for the overall population and for groups within the population: men and women, whites and blacks, teenagers and prime-age workers.

Figure 2-4 shows the breakdown of the population into the three categories for June 2017. The statistics broke down as follows:

Labor Force=153.2+7.0=160.2 million.Unemployment Rate=(7.0/160.2)×100=4.4%Labor-Force Participation Rate=(160.2/255.0)×100=62.8%.

$$\begin{aligned} \text{Labor Force} &= 153.2 + 7.0 = 160.2 \text{ million.} \\ \text{Unemployment Rate} &= (7.0/160.2) \times 100 = 4.4\% \\ \text{Labor-Force Participation Rate} &= (160.2/255.0) \times 100 = 62.8\%. \end{aligned}$$

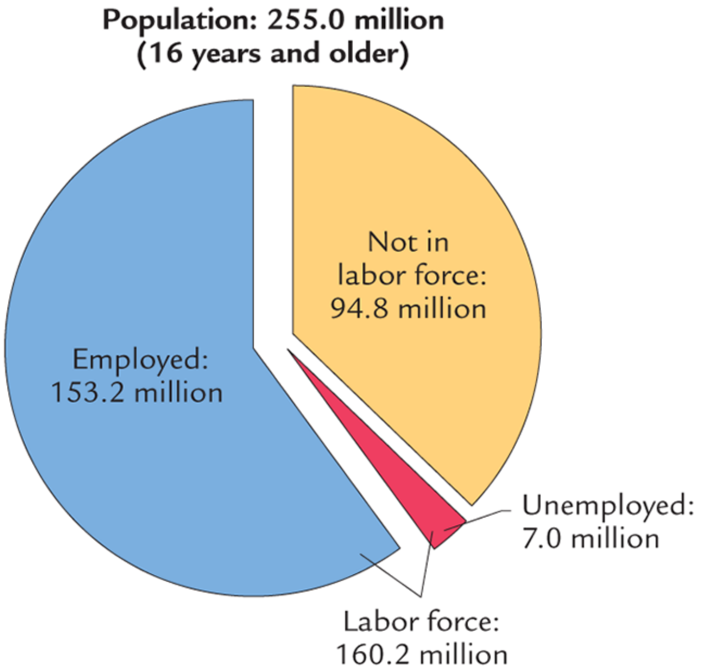


FIGURE 2-4 The Three Groups of the Population When the Bureau of Labor Statistics surveys the population, it

places all adults into one of three categories: employed, unemployed, or not in the labor force. This figure shows the number of people in each category in June 2017.

Data from: U.S. Department of Labor.

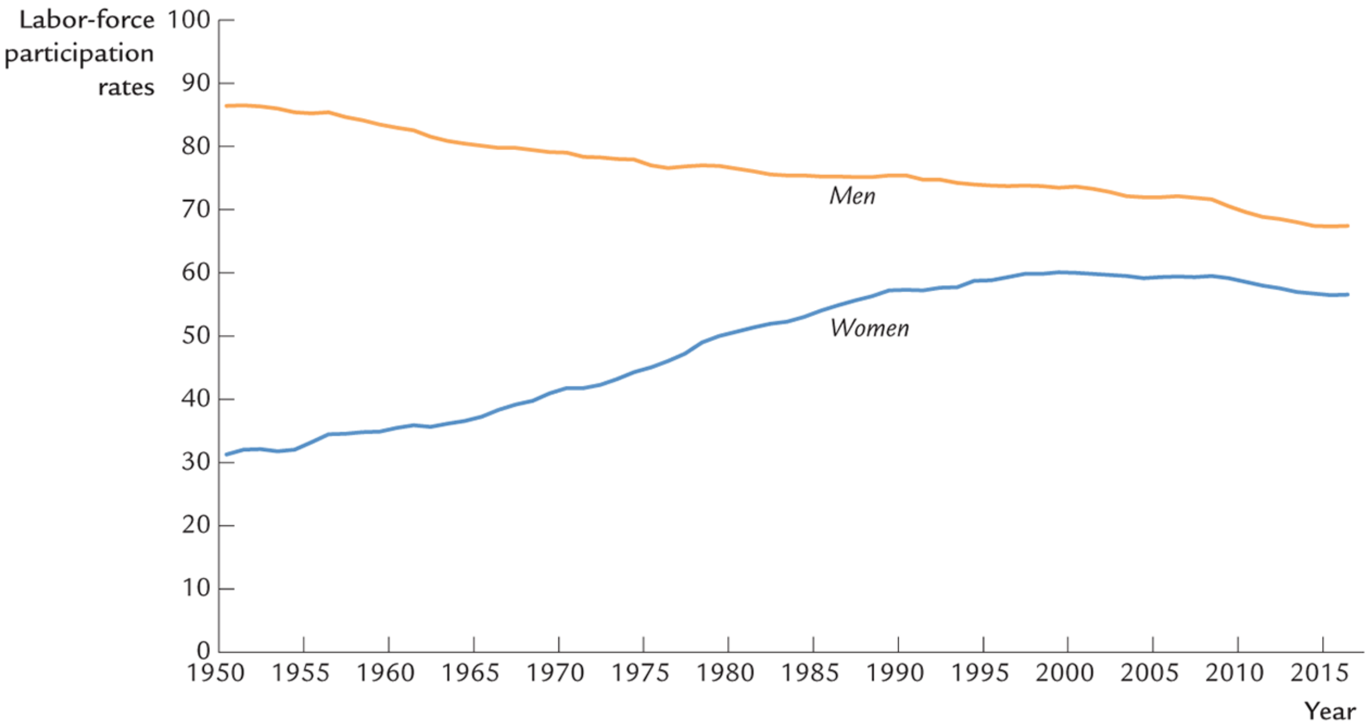
Hence, almost two-thirds of the adult population was in the labor force and 4.4 percent of those in the labor force did not have a job.

CASE STUDY

Men, Women, and Labor-Force Participation

The data on the labor market collected by the BLS reflect not only economic developments, such as the booms and busts of the business cycle, but also a variety of social changes. Longer-term social changes in the roles of men and women in society, for example, are evident in the data on labor-force participation.

Figure 2-5 shows the labor-force participation rates of men and women in the United States from 1950 to 2016. Just after World War II, men and women had very different economic roles. Only 34 percent of women were working or looking for work, in contrast to 86 percent of men. Since then, the difference between the participation rates of men and women has gradually diminished, as growing numbers of women have entered the labor force and some men have left it. Data for 2016 show that more than 56 percent of women were in the labor force, in contrast to 69 percent of men. As measured by labor-force participation, men and women are now playing more equal roles in the economy.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 2-5 Labor-Force Participation Over the past several decades, the labor-force participation rate for women has risen, while the rate for men has declined.

Data from: U.S. Department of Labor.

There are many reasons for this change. In part, it is due to new technologies, such as the washing machine, clothes dryer, refrigerator, freezer, and dishwasher, which have reduced the amount of time required to complete routine household tasks. In part, it is due to improved birth control, which has reduced the number of children born to the typical family. And in part, this change in women's role is due to changing political and social attitudes. Together, these developments have had a profound impact, as demonstrated by these data.

Although the increase in women's labor-force participation is easily explained, the fall in men's participation may seem puzzling. There are several developments at work. First, young men now stay in school longer than their fathers and grandfathers did. Second, older men now retire earlier and live longer. Third, with more women employed, more fathers now stay at home to raise their children. Full-time students, retirees, and stay-at-home fathers are all counted as being out of the labor force.

Figure 2-5 shows that, in the most recent decade, the labor-force participation rate declined for both men and women. This phenomenon is examined in Chapter 7. We will see that much of this recent decline is due to the start of retirement for the large baby-boom generation. ■

The Establishment Survey

When the BLS reports the unemployment rate every month, it also reports a variety of other statistics describing conditions in the labor market. Some of these statistics, such as the labor-force participation rate, are derived from the Current Population Survey. Other statistics come from a separate survey of about 160,000 business establishments that employ more than 40 million workers. When you read a headline that says the economy created a certain number of jobs last month, that statistic is the change in the number of workers that businesses report having on their payrolls.

Because the BLS conducts two surveys of labor-market conditions, it produces two measures of total employment. From the household survey, it obtains an estimate of the number of people who say they are working. From the establishment survey, it obtains an estimate of the number of workers firms have on their payrolls.

One might expect these two measures of employment to be identical, but that is not the case. Although they are positively correlated, the two measures can diverge, especially over short periods of time. An example of a large divergence occurred in the early 2000s, as the economy recovered from the recession of 2001. From November 2001 to August 2003, the establishment survey showed a decline in employment of 1.0 million, while the household survey showed an increase of 1.4 million. Some commentators said the economy was experiencing a "jobless recovery," but this description applied only to the establishment data, not to the household data.

Why might these two measures of employment diverge? Part of the explanation is that the surveys measure different things. For example, a person who runs his or her own business is self-employed. The household survey counts that person as working, whereas the establishment survey does not because that person does not show up on any firm's payroll. As another example, a person who holds two jobs is counted as one employed person in the household survey but is counted twice in the establishment survey because that person would show up on the payrolls of two firms.

In addition, the two measures of employment diverge because the surveys are imperfect. For example, when new firms start up, it may take some time before those firms are included in the establishment survey. The BLS tries to estimate employment at start-ups, but the model it uses to produce these estimates is a possible source of error. A different problem arises from how the household survey extrapolates employment among the surveyed households to the entire population. If the BLS uses incorrect estimates of the size of the population, these errors will be reflected in its estimates of household employment. One possible source of incorrect population estimates is changes in the rate of immigration, both legal and illegal.

In the end, the divergence between the household and establishment surveys from 2001 to 2003 remains a mystery. Some economists believe that the establishment survey is the more accurate one because it has a larger sample. One study suggests that the best measure of employment is an average of the two surveys.[6]

More important than the specifics of these surveys or this particular episode when they diverged is the broader lesson: all economic statistics are imperfect. They offer valuable information about what is happening in the economy, but each should be interpreted with a healthy dose of caution.

2-4 Conclusion: From Economic Statistics to Economic Models

The three statistics discussed in this chapter — gross domestic product, the consumer price index, and the unemployment rate — quantify the performance of the economy. Public and private decisionmakers use these statistics to monitor changes in the economy and to formulate appropriate policies. Economists use these statistics to develop and test theories about how the economy works.

In the chapters that follow, we examine some of these theories. That is, we build models that explain how these variables are determined and how economic policy affects them. Having learned how to measure economic performance, we are now ready to learn how to explain it.

CHAPTER 3

National Income: Where It Comes From and Where It Goes

Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

> A large income is the best recipe for happiness I ever heard of.
>
> —Jane Austen

The most important macroeconomic variable is gross domestic product (GDP). As we have seen, GDP measures both a nation's total output of goods and services and its total income. To appreciate the significance of GDP, one need only take a quick look at international data: compared with their poorer counterparts, nations with a high level of GDP per person have everything from better childhood nutrition to more computers per household. A large GDP does not ensure that all of a nation's citizens are happy, but it may be the best recipe for happiness that macroeconomists have to offer.

This chapter addresses four groups of questions about the sources and uses of a nation's GDP:

- How much do the firms in the economy produce? What determines a nation's total income?
- Who gets the income from production? How much goes to compensate workers, and how much goes to compensate owners of capital?
- Who buys the output of the economy? How much do households purchase for consumption, how much do households and firms purchase for investment, and how much does the government buy for public purposes?
- What equilibrates the demand for and supply of goods and services? What ensures that desired spending on consumption, investment, and government purchases equals the level of production?

To answer these questions, we must examine how the various parts of the economy interact.

A good place to start is the circular flow diagram. In Chapter 2, we traced the circular flow of dollars in a hypothetical economy that used one input (labor services) to produce one output (bread). Figure 3-1 more accurately reflects how real economies function. It shows the linkages among the economic actors—

households, firms, and the government—and how dollars flow among them through the various markets in the economy.

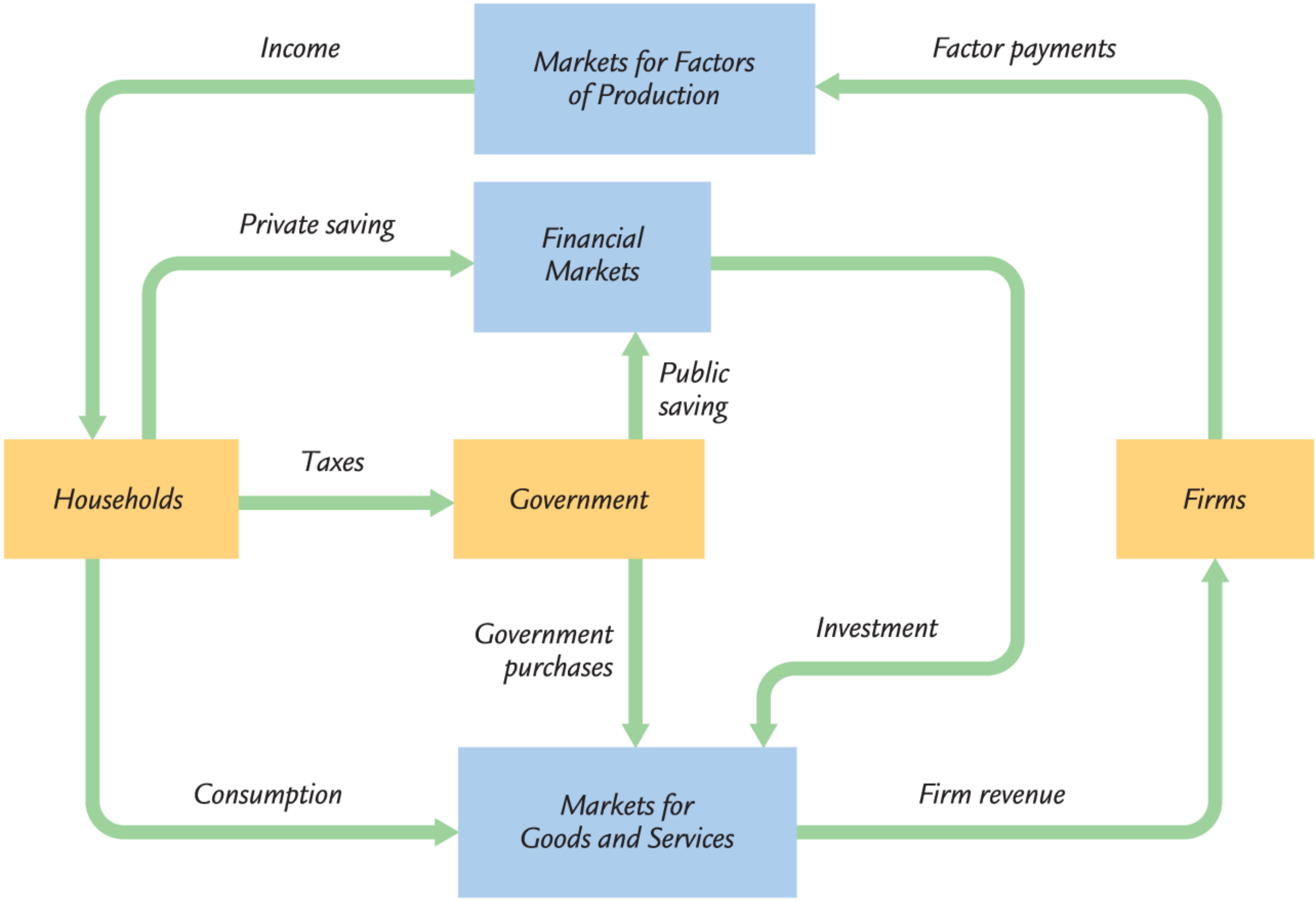


FIGURE 3-1 The Circular Flow of Dollars Through the Economy This figure is a more realistic version of the circular flow diagram in Chapter 2. Each yellow box represents an economic actor—households, firms, and the government. Each blue box represents a type of market—the markets for goods and services, the markets for the factors of production, and the financial markets. The green arrows show the flow of dollars among the economic actors through the three types of markets.

Let's look at the flow of dollars from the viewpoints of these actors. Households receive income and use it to pay taxes to the government, to consume goods and services, and to save through the financial markets. Firms receive revenue from the sale of the goods and services they produce and use it to pay for the factors of production. Households and firms borrow in financial markets to buy investment goods, such as houses and factories. The government receives revenue from taxes and uses it to pay for government purchases. Any excess of tax revenue over government spending is called *public saving*, which can be either positive (a *budget surplus*) or negative (a *budget deficit*).

In this chapter we develop a basic classical model to explain the economic interactions depicted in Figure 3-1. We begin with firms and look at what determines their level of production (and thus the level of national income). Then we examine how the markets for the factors of production distribute this income to households. Next, we consider how much of this income households consume and how much they save. In addition to discussing the demand for goods and services arising from the consumption of households, we discuss the demand arising from investment and government purchases. Finally, we come full circle and examine how the demand for goods and services (the sum of consumption, investment, and government purchases) and the supply of goods and services (the level of production) are brought into balance.

3-1 What Determines the Total Production of Goods and Services?

An economy's output of goods and services—its GDP—depends on (1) its quantity of inputs, called the factors of production, and (2) its ability to turn inputs into output, as represented by the production function.

The Factors of Production

Factors of production are the inputs used to produce goods and services. The two most important factors of production are capital and labor. *Capital* is the set of tools that workers use: the construction worker's crane, the accountant's calculator, and this author's personal computer. *Labor* is the time people spend working. We use the symbol *K* to denote the amount of capital and the symbol *L* to denote the amount of labor.

In this chapter we take the economy's factors of production as given. In other words, we assume that the economy has fixed amounts of capital and labor. We write

K=K‾.L=L‾.

$$K = \overline{K}.$$
$$L = \overline{L}.$$

The overbar means that each variable is fixed at some level. In Chapter 8, we examine what happens when the factors of production change over time, as they do in the real world. For now, to keep the analysis simple, we assume fixed amounts of capital and labor.

We also assume here that the factors of production are fully utilized. That is, no resources are wasted. Again, in the real world, part of the labor force is unemployed, and some capital lies idle. In Chapter 7, we examine the reasons for unemployment, but for now we assume that capital and labor are fully employed.

The Production Function

The available production technology determines how much output is produced from given amounts of capital and labor. Economists express this relationship using a **production function**. Letting *Y* denote the amount of output, we write the production function as

$$Y = F(K,\ L).$$

This equation states that output is a function of the amounts of capital and labor.

The production function reflects the available technology for turning capital and labor into output. If someone invents a better way to produce a good, the result is more output from the same amounts of capital and labor. Thus, technological change alters the production function.

Many production functions have a property called **constant returns to scale**. A production function has constant returns to scale if an increase of an equal percentage in all factors of production causes an increase in output of the same percentage. If the production function has constant returns to scale, then we get 10 percent more output when we increase both capital and labor by 10 percent. Mathematically, a production function has constant returns to scale if

$$zY = F(zK,\ zL)$$

for any positive number *z*. This equation says that if we multiply both the amount of capital and the amount of labor by some number *z*, output is also multiplied by *z*. In the next section, we see that the assumption of constant returns to scale has an important implication for how the income from production is distributed.

As an example of a production function, consider production at a bakery. The kitchen and its equipment are the bakery's capital, the workers hired to make the bread are its labor, and the loaves of bread are its output. The bakery's production function shows that the number of loaves produced depends on the amount of equipment and the number of workers. If the production function has constant returns to scale, then doubling the amount of equipment and the number of workers doubles the amount of bread produced.

The Supply of Goods and Services

The factors of production and the production function together determine the quantity of goods and services supplied, which in turn equals the economy's output. To express this mathematically, we write

$$\begin{aligned} Y &= F(\overline{K},\ \overline{L}) \\ &= \overline{Y}. \end{aligned}$$

In this chapter, because we assume that the supplies of capital and labor and the technology are fixed, output is also fixed (at a level denoted as $\overline{Y}$). When we discuss economic growth in Chapters 8 and 9, we will examine how increases in capital and labor and advances in technology lead to growth in the economy's output.

3-2 How Is National Income Distributed to the Factors of Production?

As we discussed in Chapter 2, the total output of an economy equals its total income. Because the factors of production and the production function together determine the total output of goods and services, they also determine national income. The circular flow diagram in Figure 3-1 shows that this national income flows from firms to households through the markets for the factors of production.

In this section, we continue to develop our model of the economy by discussing how these factor markets work. Economists have long studied factor markets to understand the distribution of income. For example, Karl Marx, the noted nineteenth-century economist, spent much time trying to explain the incomes of capital and labor. The political philosophy of communism was in part based on Marx's now-discredited theory.

Here we examine the modern theory of how national income is divided among the factors of production. It is based on the classical (eighteenth-century) idea that prices adjust to balance supply and demand, applied here to the markets for the factors of production, together with the more recent (nineteenth-century) idea that the demand for each factor of production depends on the marginal productivity of that factor. This theory, called the *neoclassical theory of distribution*, is accepted by most economists today as the best place to start in understanding how the economy's income is distributed from firms to households.

Factor Prices

The distribution of national income is determined by factor prices. **Factor prices** are the amounts paid to each unit of the factors of production. In an economy where the two factors of production are capital and labor, the two factor prices are the rent the owners of capital collect and the wage workers earn.

As Figure 3-2 shows, the price each factor of production receives for its services is determined by the supply and demand for that factor. Because we have assumed that the economy's factors of production are fixed, the factor supply curve in Figure 3-2 is vertical. Regardless of the factor price, the quantity of the factor supplied to the market is the same. The intersection of the downward-sloping factor demand curve and the vertical supply curve determines the equilibrium factor price.

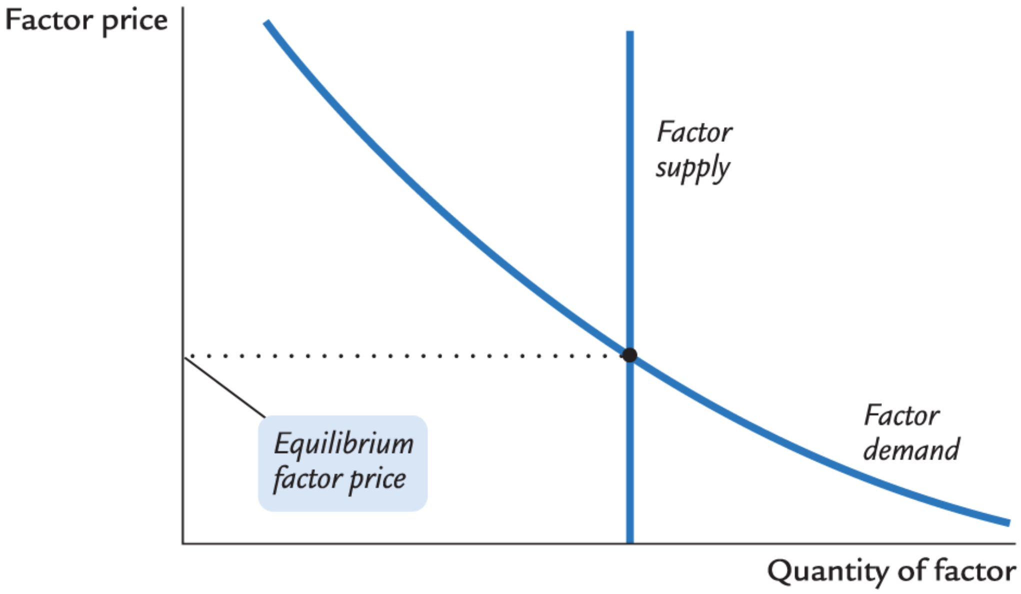


FIGURE 3-2 How a Factor of Production Is Compensated The price paid to any factor of production depends on the supply and demand for that factor's services. Because we have assumed that supply is fixed, the supply curve is vertical. The demand curve is downward sloping. The intersection of the supply and demand curves determines the equilibrium factor price.

To understand factor prices and the distribution of income, we must examine the demand for the factors of production. Because factor demand arises from the thousands of firms that use capital and labor, we start by examining the decisions a typical firm makes about how much of these factors to employ.

The Decisions Facing a Competitive Firm

The simplest assumption to make about a typical firm is that it is competitive. A **competitive firm** is small relative to the markets in which it trades, so it has little influence on market prices. For example, our firm produces a good and sells it at the market price. Because many firms produce this good, our firm can sell as much as it wants without causing the price of the good to fall, or it can stop selling altogether without causing the price of the good to rise. Similarly, our firm cannot influence the wages of the workers it employs because many other local firms also employ workers. The firm has no reason to pay more than the market wage, and if it tried to pay less, its workers would take jobs elsewhere. Therefore, the competitive firm takes the prices of its output and its inputs as given by market conditions.

To make its product, the firm needs two factors of production, capital and labor. As we did for the aggregate economy, we represent the firm's production technology with the production function

$$Y = F(K, L),$$

where *Y* is the number of units produced (the firm's output), *K* the number of machines used (the amount of capital), and *L* the number of hours worked by the firm's employees (the amount of labor). Holding constant the technology as expressed in the production function, the firm produces more output only if it uses more machines or if its employees work more hours.

The firm sells its output at a price *P*, hires workers at a wage *W*, and rents capital at a rate *R*. Notice that when we speak of firms renting capital, we are assuming that households own the economy's stock of capital. In this analysis, households rent out their capital, just as they sell their labor. The firm obtains both factors of production from the households that own them.[1]

The goal of the firm is to maximize profit. **Profit** equals revenue minus costs; it is what the owners of the firm keep after paying for the costs of production. Revenue equals $P \times Y$, the selling price of the good *P* multiplied by the amount of the good the firm produces *Y*. Costs include labor and capital costs. Labor costs equal $W \times L$, the wage *W* times the amount of labor *L*. Capital costs equal $R \times K$, the rental price of capital *R* times the amount of capital *K*. We can write

$$\begin{aligned} \text{Profit} &= \text{Revenue} - \text{Labor Costs} - \text{Capital Costs} \\ &= PY - WL - RK. \end{aligned}$$

To see how profit depends on the factors of production, we use the production function $Y = F(K, L)$ to substitute for *Y* to obtain

$$\text{Profit} = PF(K, L) - WL - RK.$$

This equation shows that profit depends on the product price *P*, the factor prices *W* and *R*, and the factor quantities *L* and *K*. The competitive firm takes the product price and the factor prices as given and chooses the amounts of labor and capital that maximize profit.

The Firm's Demand for Factors

We now know that our firm will hire labor and rent capital in the quantities that maximize profit. But what are those profit-maximizing quantities? To answer this question, we first consider the quantity of labor and then the quantity of capital.

The Marginal Product of Labor

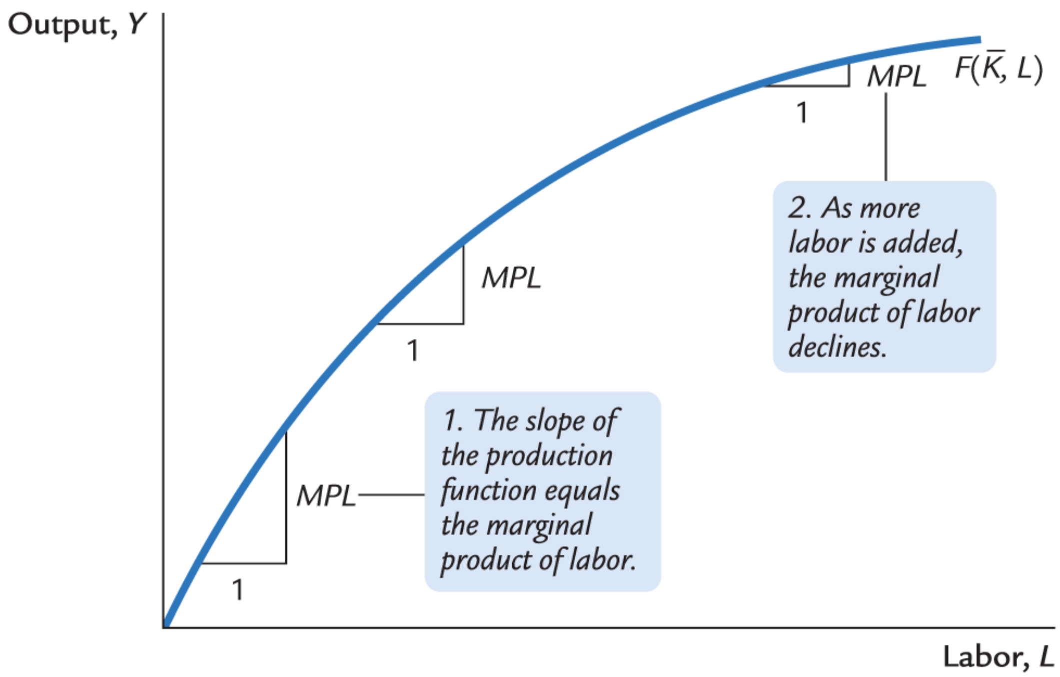
The more labor the firm employs, the more output it produces. The **marginal product of labor (*MPL*)** is the extra amount of output the firm gets from one extra unit of labor, holding the amount of capital fixed. We can express this using the production function:

$$MPL = F(K,\ L+1) - F(K,\ L).$$

The first term on the right-hand side is the amount of output produced with *K* units of capital and $L + 1$ units of labor; the second term is the amount of output produced with *K* units of capital and *L* units of labor. This equation states that the marginal product of labor is the difference between the amount of output produced with $L + 1$ units of labor and the amount produced with only *L* units of labor.

Most production functions have the property of **diminishing marginal product**: holding the amount of capital fixed, the marginal product of labor decreases as the amount of labor increases. To see why, consider again the production of bread at a bakery. As a bakery hires more labor, it produces more bread. The *MPL* is the amount of extra bread produced when an extra unit of labor is hired. As more labor is added to a fixed amount of capital, however, the *MPL* falls. Fewer additional loaves are produced because workers are less productive when the kitchen is more crowded. In other words, holding the size of the kitchen fixed, each additional worker adds fewer loaves of bread to the bakery’s output.

Figure 3-3 graphs the production function. It illustrates what happens to the amount of output when we hold the amount of capital constant and vary the amount of labor. This figure shows that the marginal product of labor is the slope of the production function. As the amount of labor increases, the production function becomes flatter, indicating diminishing marginal product.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 3-3 The Production Function This curve shows how output depends on labor input, holding the amount of capital constant. The marginal product of labor *MPL* is the change in output when the labor input is increased by 1 unit. As the amount of labor increases, the production function becomes flatter, indicating diminishing marginal product.

From the Marginal Product of Labor to Labor Demand

When the competitive, profit-maximizing firm is deciding whether to hire an additional unit of labor, it considers how that decision would affect profits. It therefore compares the extra revenue from increased production with the extra cost from hiring the additional labor. The increase in revenue from an additional unit of labor depends on two variables: the marginal product of labor and the price of the output. Because an extra unit of labor produces *MPL* units of output and each unit of output sells for *P* dollars, the extra revenue is $P \times MPL$. The extra cost of hiring one more unit of labor is the wage *W*. Thus, the change in profit from hiring an additional unit of labor is

$$\begin{aligned} \Delta\ \text{Profit} &= \Delta\ \text{Revenue} - \Delta\ \text{Cost} \\ &= (P \times MPL) - W. \end{aligned}$$

The symbol Δ (called *delta*) denotes the change in a variable.

We can now answer the question we asked at the beginning of this section: how much labor does the firm hire? The firm's manager knows that if the extra revenue $P \times MPL$ exceeds the wage *W*, an extra unit of labor increases profit. Therefore, the manager continues to hire labor until the next unit would no longer be profitable—that is, until the *MPL* falls to the point where the extra revenue equals the wage. The competitive firm’s demand for labor is determined by

P×MPL=W.
$$P \times MPL = W.$$

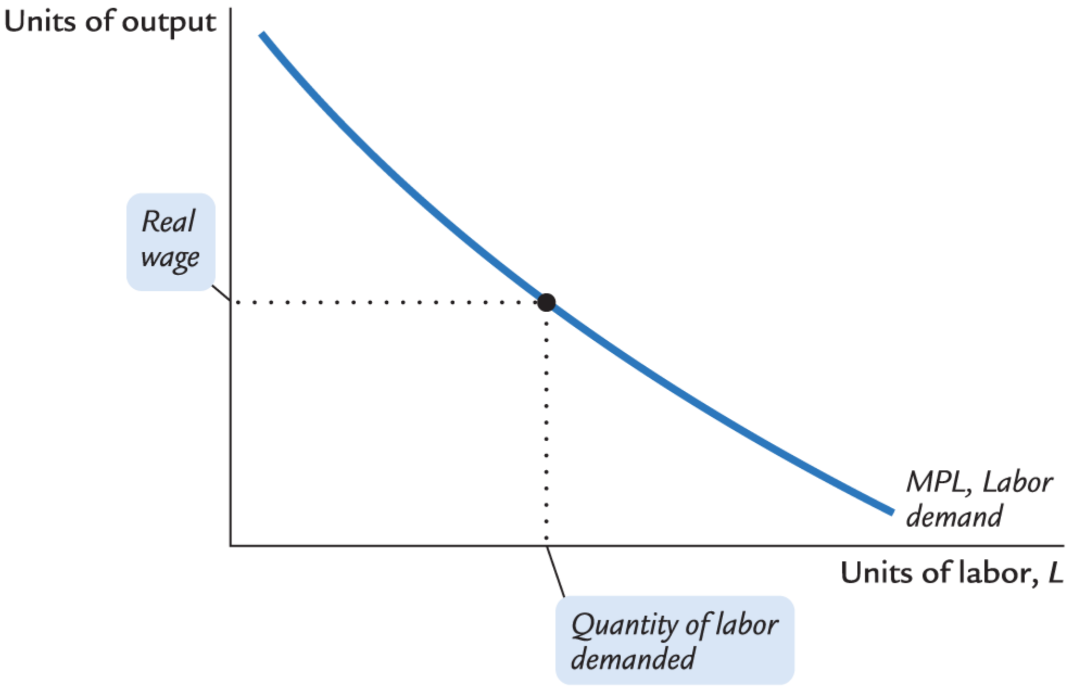
We can also write this as

MPL=W/P.
$$MPL = W/P.$$

W/P is the **real wage**—the payment to labor measured in units of output rather than in dollars. To maximize profit, the firm hires up to the point at which the marginal product of labor equals the real wage.

For example, again consider a bakery. Suppose the price of bread *P* is $2 per loaf, and a worker earns a wage *W* of $20 per hour. The real wage *W/P* is 10 loaves per hour. In this example, the firm keeps hiring workers as long as the additional worker would produce at least 10 loaves per hour. When the *MPL* falls to 10 loaves per hour or less, hiring additional workers is no longer profitable.

Figure 3-4 shows how the marginal product of labor depends on the amount of labor employed (holding the firm’s capital stock constant). That is, this figure graphs the *MPL* schedule. Because the *MPL* diminishes as the amount of labor increases, this curve slopes downward. For any given real wage, the firm hires up to the point at which the *MPL* equals the real wage. Hence, the *MPL* schedule is also the firm’s labor demand curve.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 3-4 The Marginal Product of Labor Schedule The marginal product of labor *MPL* depends on the amount of labor. The *MPL* curve slopes downward because the *MPL* declines as *L* increases. The firm hires labor up to the point where the real wage *W/P* equals the *MPL*. Hence, this schedule is also the firm’s labor demand curve.

The Marginal Product of Capital and Capital Demand

The firm decides how much capital to rent in the same way it decides how much labor to hire. The **marginal product of capital (*MPK*)** is the amount of extra output the firm gets from an extra unit of capital, holding the amount of labor constant:

$$MPK = F(K + 1,\ L) - F(K,\ L).$$

Thus, the marginal product of capital is the difference between the amount of output produced with $K + 1$ units of capital and that produced with only *K* units of capital.

Like labor, capital is subject to diminishing marginal product. Once again consider the production of bread at a bakery. The first several ovens installed in the kitchen will be very productive. However, if the bakery installs more and more ovens, while holding its labor force constant, it will eventually contain more ovens than its employees can effectively operate. Hence, the marginal product of the last few ovens is lower than that of the first few.

The increase in profit from renting an additional machine is the extra revenue from selling the output of that machine minus the machine's rental price:

$$\begin{aligned} \Delta\ \text{Profit} &= \Delta\ \text{Revenue} - \Delta\ \text{Cost} \\ &= (P \times MPK) - R. \end{aligned}$$

To maximize profit, the firm continues to rent more capital until the *MPK* falls to equal the real rental price:

$$MPK = R/P.$$

The **real rental price of capital** is the rental price measured in units of goods rather than in dollars.

To sum up, the competitive, profit-maximizing firm follows a simple rule about how much labor to hire and how much capital to rent. *The firm demands each factor of production until that factor's marginal product equals its real factor price*.

The Division of National Income

Having analyzed how a firm decides the quantity of each factor to employ, we can now explain how the markets for the factors of production distribute the economy's total income. If all firms in the economy are competitive and profit maximizing, then each factor of production is paid its marginal contribution to the production process. The real wage paid to each worker equals the *MPL*, and the real rental price paid to each owner of capital equals the *MPK*. The total real wages paid to labor are therefore $MPL \times L$, and the total real return paid to capital owners is $MPK \times K$.

The income that remains after the firms have paid the factors of production is the **economic profit** of the owners of the firms:

$$\text{Economic Profit} = Y - (MPL \times L) - (MPK \times K).$$

Note that income *Y* and economic profit are here being expressed in real terms—that is, in units of output rather than in dollars. Because we want to examine the distribution of income, we rearrange the terms as follows:

$$Y = (MPL \times L) + (MPK \times K) + \text{Economic Profit}.$$

Total income is divided among the return to labor, the return to capital, and economic profit.

How large is economic profit? The answer is surprising: if the production function has the property of constant returns to scale, as is often thought to be the case, then economic profit must be zero. That is, nothing is left after the factors of production are paid. This conclusion follows from a famous mathematical result called *Euler's theorem*,[2] which states that if the production function has constant returns to scale, then

$$F(K,\ L) = (MPK \times K) + (MPL \times L).$$

If each factor of production is paid its marginal product, then the sum of these factor payments equals total output. In other words, constant returns to scale, profit maximization, and competition together imply that economic profit is zero.

If economic profit is zero, how can we explain the existence of "profit" in the economy? The answer is that the term *profit* as normally used is different from economic profit. We have been assuming that there are three types of agents: workers, owners of capital, and owners of firms. Total income is divided among wages, return to capital, and economic profit. In the real world, however, most firms own rather than rent the capital they use. Because firm owners and capital owners are the same people, economic profit and the return to capital are often lumped together. If we call this alternative definition **accounting profit**, we can say that

Accounting Profit=Economic Profit+(MPK×K).

$$\text{Accounting Profit} = \text{Economic Profit} + (MPK \times K).$$

Under our assumptions—constant returns to scale, profit maximization, and competition—economic profit is zero. If these assumptions approximately describe the world, then the "profit" in the national income accounts must be mostly the return to capital.

We can now answer the question posed at the beginning of this chapter about how the income of the economy is distributed from firms to households. Each factor of production is paid its marginal product, and these factor payments exhaust total output. *Total output is divided between the payments to capital and the payments to labor, depending on their marginal productivities*.

CASE STUDY

The Black Death and Factor Prices

According to the neoclassical theory of distribution, factor prices equal the marginal products of the factors of production. Because the marginal products depend on the quantities of the factors, a change in the quantity of any one factor alters the marginal products of all the factors. Therefore, a change in the supply of a factor alters equilibrium factor prices and the distribution of income.

Fourteenth-century Europe provides a grisly natural experiment to study how factor quantities affect factor prices. The outbreak of the bubonic plague—the Black Death—in 1348 reduced the population of Europe by about one-third within a few years. Because the marginal product of labor increases as the amount of labor falls, this massive reduction in the labor force should have raised the marginal product of labor and equilibrium real wages. (That is, the economy should have moved to the left along the curves in Figures 3-3 and 3-4.) The evidence confirms the theory: real wages approximately doubled during the plague years. The peasants who were fortunate enough to survive the plague enjoyed economic prosperity.

The reduction in the labor force caused by the plague should also have affected the return to land, the other major factor of production in medieval Europe. With fewer workers available to farm the land, an additional unit of land would have produced less additional output, and so land rents should have fallen. Once again, the theory is confirmed: real rents fell 50 percent or more during this period. While the peasant classes prospered, the landed classes suffered reduced incomes.[3] ■

The Cobb–Douglas Production Function

What production function describes how actual economies turn capital and labor into GDP? One answer to this question came from a historic collaboration between a U.S. senator and a mathematician.

Paul Douglas was a U.S. senator from Illinois from 1949 to 1967. In 1927, however, when he was still a professor of economics, he noticed a surprising fact: the division of national income between capital and labor had been roughly constant over a long period. In other words, as the economy grew more prosperous over time, the total income of workers and the total income of capital owners grew at almost exactly the same rate. This observation caused Douglas to wonder what conditions might lead to constant factor shares.

Douglas asked Charles Cobb, a mathematician, what production function, if any, would produce constant factor shares if factors always earned their marginal products. The production function would need to have the property that

$$\text{Capital Income} = MPK \times K = \alpha Y$$

and

$$\text{Labor Income} = MPL \times L = (1 - \alpha)Y,$$

where α is a constant between zero and one that measures capital's share of income. That is, α determines what share of income goes to capital and what share goes to labor. Cobb showed that the function with this property is

$$F(K,\ L) = AK^{\alpha}L^{1-\alpha},$$

where *A* is a parameter greater than zero that measures the productivity of the available technology. This function became known as the **Cobb–Douglas production function**.

Let's take a closer look at some of the properties of this production function. First, the Cobb–Douglas production function has constant returns to scale. That is, if capital and labor are increased by the same proportion, then output increases by that proportion as well.[4]

Next, consider the marginal products for the Cobb–Douglas production function. The marginal product of labor is[5]

$$MPL = (1 - \alpha) AK^{\alpha} L^{-\alpha},$$

and the marginal product of capital is

$$MPK = \alpha AK^{\alpha - 1} L^{1-\alpha}.$$

From these equations, recalling that α is between zero and one, we can see what causes the marginal products of the two factors to change. An increase in the amount of capital raises the *MPL* and reduces the *MPK*. Similarly, an increase in the amount of labor reduces the *MPL* and raises the *MPK*. A technological advance that increases the parameter *A* raises the marginal product of both factors proportionately.

The marginal products for the Cobb–Douglas production function can also be written as[6]

$$MPL = (1 - \alpha) Y / L.$$
$$MPK = \alpha Y / K.$$

The *MPL* is proportional to output per worker, and the *MPK* is proportional to output per unit of capital. *Y/L* is called *average labor productivity*, and *Y/K* is called *average capital productivity*. If the production function is Cobb–Douglas, then the marginal productivity of a factor is proportional to its average productivity.

We can now verify that if factors earn their marginal products, then the parameter α indeed tells us how much income goes to labor and how much goes to capital. The total amount paid to labor, which we have seen is $MPL \times L$, equals $(1 - \alpha)Y$. Therefore, $(1 - \alpha)$ is labor's share of output. Similarly, the total amount paid to capital, $MPK \times K$, equals αY, and α is capital's share of output. The ratio of labor income to capital income is a constant, $(1 - \alpha)/\alpha$, just as Douglas observed. The factor shares depend only on the parameter α, not on the amounts of capital or labor or on the state of technology as measured by the parameter *A*.

More recent U.S. data are also consistent with the Cobb–Douglas production function. Figure 3-5 shows the ratio of labor income to total income in the United States from 1960 to 2016. Despite the many changes in the economy over the past five decades, this ratio has remained about 2/3. This division of income is easily explained by a Cobb–Douglas production function in which the parameter α is about 1/3. According to this parameter, capital receives one-third of income, and labor receives two-thirds.

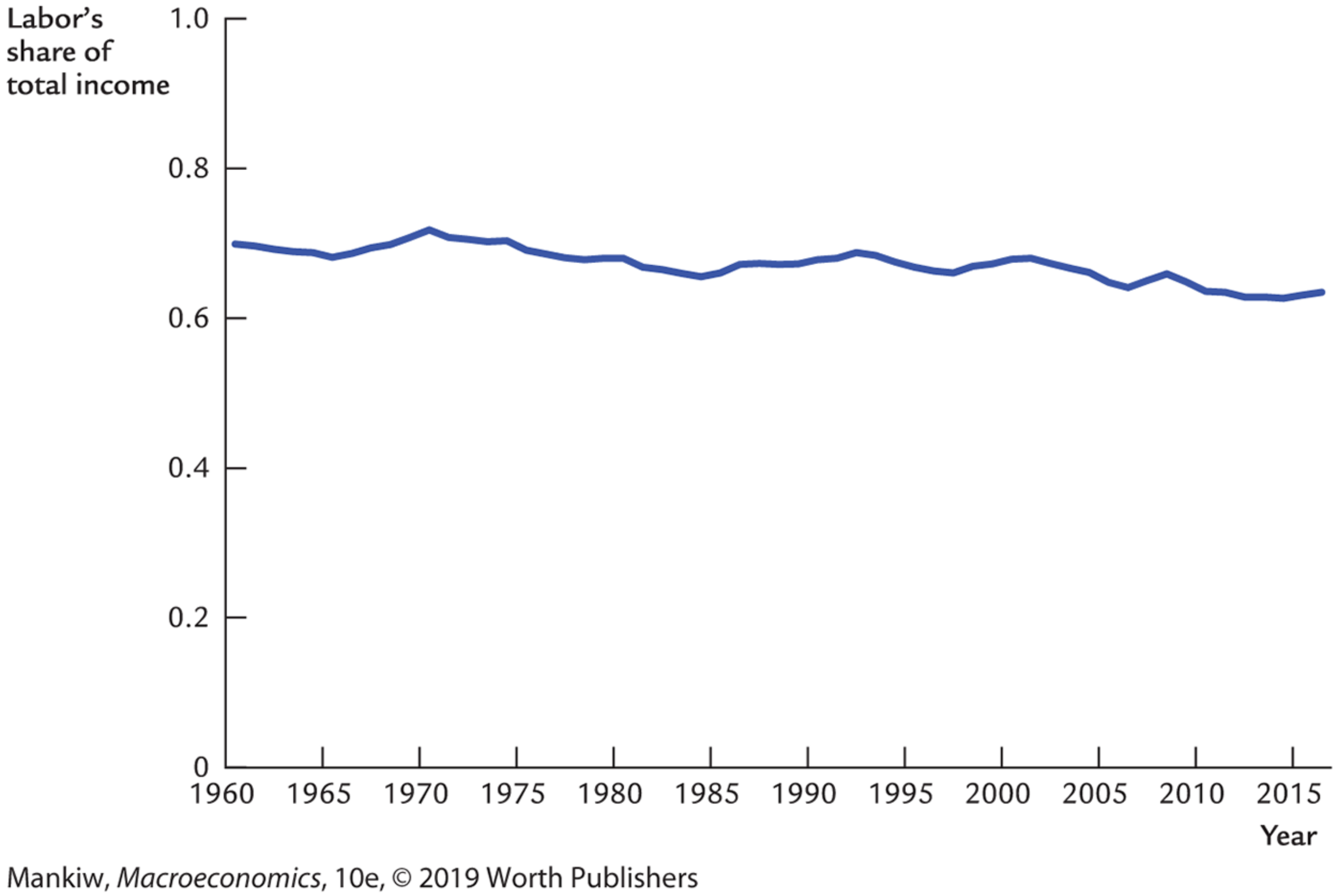


FIGURE 3-5 The Ratio of Labor Income to Total Income Labor income has remained about two-thirds of total income over a long period of time. This approximate constancy of factor shares is consistent with the Cobb–Douglas production function.

Data from: U.S. Department of Commerce. This figure is produced from U.S. national income accounts data. Labor income is compensation of employees. Total income is the sum of labor income, corporate profits, net interest, rental income, and depreciation. Proprietors' income is excluded from these calculations because it is a combination of labor income and capital income.

Although the capital and labor shares are approximately constant, they are not exactly constant. In Figure 3-5, the labor share fell from a high of 72 percent in 1970 to a low of 63 percent in 2014. And, of course, the capital share increased from 28 percent to 37 percent. The reason for this change is not well understood. One possibility is that technological progress over the past several decades has not simply increased the parameter *A* but may have also changed the relative importance of capital and labor in the production process, thereby altering the parameter α. Or there may be important determinants of incomes that are not well captured by the Cobb–Douglas production function together with the model of competitive product and factor markets, such as the changing market power of firms or unions.

The Cobb–Douglas production function is not the last word in explaining the economy's production of goods and services or the distribution of national income between capital and labor. It is, however, a good place to start.

CASE STUDY

Labor Productivity as the Key Determinant of Real Wages

The neoclassical theory of distribution tells us that the real wage *W/P* equals the marginal product of labor. The Cobb–Douglas production function tells us that the marginal product of labor is proportional to average labor productivity *Y/L*. If this theory is right, then workers should enjoy rapidly rising living standards when labor productivity is growing robustly. Is this true?

Table 3-1 presents some data on growth in productivity and real wages for the U.S. economy. From 1960 to 2016, productivity as measured by output per hour of work grew about 2.0 percent per year. Real wages grew at 1.8 percent—almost the same rate. With a growth rate of 2 percent per year, productivity and real wages double about every 35 years.

TABLE 3-1 Growth in Labor Productivity and Real Wages: The U.S. Experience

Time Period	Growth Rate of Labor Productivity	Growth Rate of Real Wages
1960–2016	2.0%	1.8%
1960–1973	3.0	2.7
1973–1995	1.5	1.2
1995–2010	2.6	2.2
2010–2016	0.5	0.9

Data from: U.S. Department of Labor. Growth in labor productivity is measured here as the annualized rate of change in output per hour in the nonfarm business sector. Growth in real wages is measured as the annualized change in compensation per hour in the nonfarm business sector divided by the implicit price deflator for that sector.

Productivity growth varies over time. The table shows the data for four shorter periods that economists have identified as having different productivity experiences. Around 1973, the U.S. economy experienced a significant slowdown in productivity growth that lasted until 1995. The cause of the productivity slowdown is not well understood, but the link between productivity and real wages was exactly as standard theory predicts. The slowdown in productivity growth from 3.0 to 1.5 percent per year coincided with a slowdown in real wage growth from 2.7 to 1.2 percent per year.

Productivity growth picked up again around 1995, and many observers hailed the arrival of the “new economy.” This productivity acceleration is often attributed to the spread of computers and information technology. As theory predicts, growth in real wages picked up as well. From 1995 to 2010, productivity grew by 2.6 percent per year and real wages by 2.2 percent per year. After 2010, productivity and real wages slowed down again, and commentators lamented this “new normal.” From 2010 to 2016, productivity and real wage grew less than 1 percent per year.

These changes in productivity growth are largely unpredictable, and even with the benefit of hindsight, they have proven difficult to explain. Yet theory and history both confirm the close link between labor productivity and real wages. This lesson is the key to understanding why workers today are better off than workers in previous generations. ■

FYI

The Growing Gap Between Rich and Poor

One striking change in the U.S. economy, as well as in many other economies around the world, is the increase in income inequality since the 1970s. This development is not primarily about the distribution of national income between capital and labor. Instead, it concerns the distribution of labor income between skilled workers (those with a college degree, for example) and unskilled workers (those without a college degree). Skilled workers have always been paid more than unskilled workers, but over the past several decades, the wages of skilled workers have grown more quickly than the wages of unskilled workers, exacerbating inequality.

Why has this occurred? One diagnosis comes from economists Claudia Goldin and Lawrence Katz in their book *The Race Between Education and Technology*.[7] Their bottom line is that “the sharp rise in inequality was largely due to an educational slowdown.”

According to Goldin and Katz, for the past century technological progress has been a steady economic force, not only increasing average living standards but also increasing the demand for skilled workers relative to unskilled workers. Skilled workers are needed to apply and manage new technologies, while unskilled workers are more likely to be made obsolete. (Think about robots, for instance, or even your bank’s ATM.) By itself, this *skill-biased technological change* tends to raise the wages of skilled workers relative to the wages of unskilled workers, thereby increasing inequality.

For much of the twentieth century, however, skill-biased technological change was outpaced by advances in educational attainment. In other words, while technological progress increased the demand for skilled workers, the educational system increased the supply of skilled workers even faster. As a result, skilled workers did not benefit disproportionately from economic growth. Indeed, until the 1970s, wages for skilled workers grew more slowly than wages for unskilled workers, reducing inequality.

Recently things have changed. Over the past several decades, Goldin and Katz argue, skill-biased technological change has continued, but educational advancement has slowed. The cohort of workers born in 1950 averaged 4.67 more years of schooling than the cohort born in 1900, representing an increase of 0.93 years of schooling per decade. By contrast, the cohort born in 1975 had only 0.74 more years of schooling than the one born in 1950, an increase of only 0.30 years per decade. That is, the pace of educational advancement fell by 68 percent. Because growth in the supply of skilled workers has slowed, their wages have grown relative to those of the unskilled. (Implication for personal decisionmaking: for most people, college and graduate school are investments well worth making.)

Increasing income inequality is a prominent topic in public policy debates. Some policymakers advocate a more redistributive system of taxes and transfers, to take from those higher on the economic ladder and give to those on the lower rungs. This approach treats the symptoms but not the underlying causes of rising inequality. If Goldin and Katz are correct, reversing the rise in income inequality will require putting more of society’s resources into education (which economists call *human capital*). Educational reform is a topic beyond the scope of this book, but it is worth noting that, if successful, such reform could profoundly affect the economy and the distribution of income.

3-3 What Determines the Demand for Goods and Services?

We have seen what determines the level of production and how the income from production is distributed to workers and owners of capital. We now continue our tour of the circular flow diagram, Figure 3-1, and examine how the output from production is used.

In Chapter 2, we identified the four components of GDP:

- Consumption (*C*)
- Investment (*I*)
- Government purchases (*G*)
- Net exports (*NX*).

The circular flow diagram contains only the first three components. For now, to simplify the analysis, we assume our economy is a *closed economy*—a country that does not trade with other countries. Thus, net exports are always zero. (We examine the macroeconomics of *open economies* in Chapter 6.)

A closed economy has three uses for the goods and services it produces. These three components of GDP are expressed in the *national income accounts identity*:

$$Y = C + I + G.$$

Households consume some of the economy’s output, firms and households use some of the output for investment, and the government buys some of the output for public purposes. We want to see how GDP is allocated among these three uses.

Consumption

When we eat food, wear clothing, or go to a movie, we are consuming some of the output of the economy. All forms of consumption together make up about two-thirds of GDP. Because consumption is so large, macroeconomists have devoted much energy to studying how households make their consumption decisions. Chapter 19 examines this topic in detail. Here we consider the simplest story of consumer behavior.

Households receive income from their labor and their ownership of capital, pay taxes to the government, and then decide how much of their after-tax income to consume and how much to save. As we discussed in Section 3-2, the income that households receive equals the output of the economy *Y*. The government then taxes households an amount *T*. (Although the government imposes many kinds of taxes, such as personal and corporate income taxes and sales taxes, for our purposes, we can lump all these taxes together.) We define income after the payment of all taxes, *Y* – *T*, to be **disposable income**. Households divide their disposable income between consumption and saving.

We assume that the level of consumption depends directly on the level of disposable income. A higher level of disposable income leads to greater consumption. Thus,

C=C(Y−T). $$C = C(Y - T).$$

This equation states that consumption is a function of disposable income. The relationship between consumption and disposable income is called the **consumption function**.

The **marginal propensity to consume (*MPC*)** is the amount by which consumption changes when disposable income increases by one dollar. The *MPC* is between zero and one: an extra dollar of income increases consumption but by less than one dollar. Thus, if households obtain an extra dollar of income, they save a portion of it. For example, if the *MPC* is 0.7, then households spend 70 cents of each additional dollar of disposable income on consumer goods and services and save 30 cents.

Figure 3-6 depicts the consumption function. The slope of the consumption function tells us how much consumption increases when disposable income increases by one dollar. That is, the slope of the consumption function is the *MPC*.

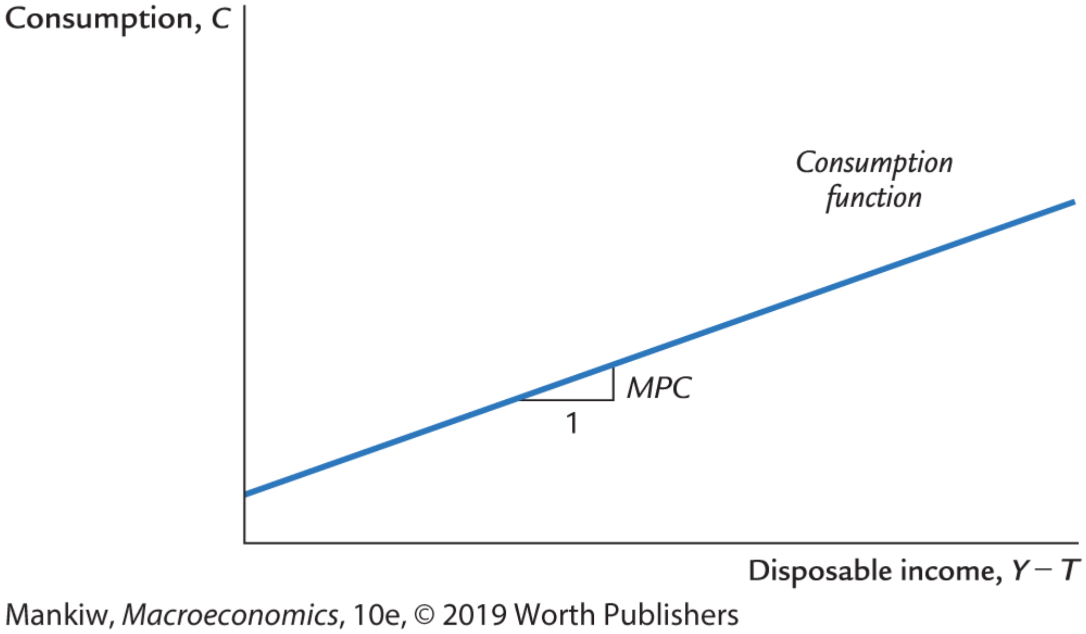


FIGURE 3-6 The Consumption Function The consumption function relates consumption *C* to disposable income *Y* – *T*. The marginal propensity to consume *MPC* is the amount by which consumption increases when disposable income

increases by one dollar.

Investment

Both firms and households purchase investment goods. Firms buy investment goods to add to their stock of capital and to replace existing capital as it wears out. Households buy new houses, which are also part of investment. Total investment in the United States averages about 15 percent of GDP.

The quantity of investment goods demanded depends on the **interest rate**, which measures the cost of the funds used to finance investment. For an investment project to be profitable, its return (the revenue from increased future production of goods and services) must exceed its cost (the payments for borrowed funds). If the interest rate rises, fewer investment projects are profitable, and the quantity of investment goods demanded falls.

For example, suppose a firm is considering whether it should build a $1 million factory that would yield a return of $100,000 per year, or 10 percent. The firm compares this return to the cost of borrowing the $1 million. If the interest rate is below 10 percent, the firm borrows the money in financial markets and makes the investment. If the interest rate is above 10 percent, the firm forgoes the investment opportunity and does not build the factory.

The firm makes the same investment decision even if it does not have to borrow the $1 million but rather uses its own funds. The firm can always deposit this money in a bank or a money market fund and earn interest on it. Building the factory is more profitable than depositing the money if and only if the interest rate is less than the 10 percent return on the factory.

A person wanting to buy a new house faces a similar decision. The higher the interest rate, the greater the cost of carrying a mortgage. A $100,000 mortgage costs $6,000 per year if the interest rate is 6 percent and $8,000 per year if the interest rate is 8 percent. As the interest rate rises, the cost of owning a home rises, and the demand for new homes falls.

When studying the role of interest rates in the economy, economists distinguish between the nominal interest rate and the real interest rate. This distinction is relevant when the overall level of prices is changing. The **nominal interest rate** is the interest rate as usually reported: it is the rate of interest that investors pay to borrow money. The **real interest rate** is the nominal interest rate corrected for the effects of inflation. If the nominal interest rate is 8 percent and the inflation rate is 3 percent, then the real interest rate is 5 percent. In Chapter 5, we discuss the relation between nominal and real interest rates in detail. Here it is sufficient to note that the real interest rate measures the true cost of borrowing and, thus, determines the quantity of investment.

We can summarize this discussion with an equation relating investment I to the real interest rate r:

$$I = I(r).$$

Figure 3-7 shows this investment function. It slopes downward because as the interest rate rises, the quantity of investment demanded falls.

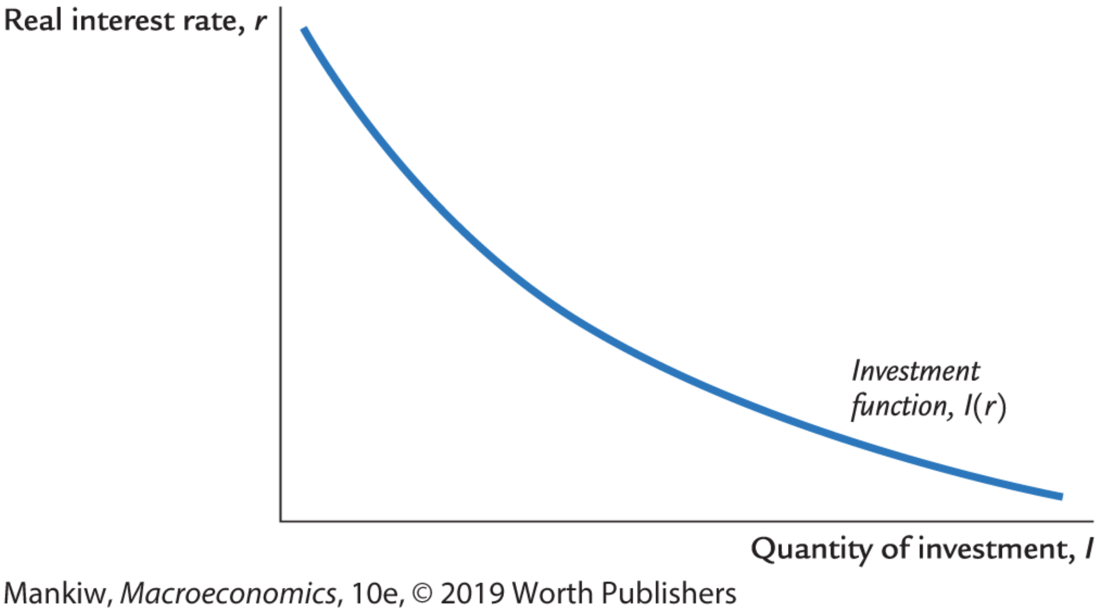


FIGURE 3-7 The Investment Function The investment function relates the quantity of investment *I* to the real interest rate *r*. Investment depends on the real interest rate because the interest rate is the cost of borrowing. The investment function slopes downward: when the interest rate rises, fewer investment projects are profitable.

FYI

The Many Different Interest Rates

If you look in the business section of a newspaper or on a financial website, you will find many different interest rates reported. By contrast, throughout this book, we talk about “the” interest rate, as if there were only one interest rate in the economy. The only distinction we make is between the nominal interest rate (which is not corrected for inflation) and the real interest rate (which is corrected for inflation). Almost all of the interest rates reported by financial news organizations are nominal.

Why are there so many interest rates? The various interest rates differ in three ways:

- *Term*. Some loans in the economy are for short periods of time, even as short as overnight. Other loans are for thirty years or even longer. The interest rate on a loan depends on its term. Long-term interest rates are usually, but not always, higher than short-term interest rates.
- *Credit risk*. In deciding whether to make a loan, a lender must take into account the probability that the borrower will repay. The law allows borrowers to default on their loans by declaring bankruptcy. The higher the perceived probability of default, the higher the interest rate. Because the government has the lowest credit risk, government bonds tend to pay a low interest rate. At the other extreme, financially shaky corporations can raise funds only by issuing *junk bonds*, which pay a high interest rate to compensate for the high risk of default.
- *Tax treatment*. The interest on different types of bonds is taxed differently. Most important, when state and

local governments issue bonds, called *municipal bonds*, the holders of the bonds do not pay federal income tax on the interest income. Because of this tax advantage, municipal bonds pay a lower interest rate.

When you see two different interest rates reported, you can almost always explain the difference by considering the term, the credit risk, and the tax treatment of the loan.

Although there are many different interest rates in the economy, macroeconomists often ignore these distinctions because the various interest rates tend to rise and fall together. For many purposes, we will not go far wrong by assuming there is only one interest rate.

Government Purchases

Government purchases are the third component of the demand for goods and services. The federal government buys guns, missiles, and the services of government employees. Local governments buy library books, build schools, and hire teachers. Governments at all levels build roads and other public works. All these transactions make up government purchases of goods and services, which account for about 20 percent of GDP in the United States.

These purchases are only one type of government spending. The other is transfer payments to households, such as public assistance for the poor and Social Security payments for the elderly. Unlike government purchases, transfer payments are not made in exchange for some of the economy's output of goods and services. Therefore, they are not included in the variable G.

Transfer payments do affect the demand for goods and services indirectly. Transfer payments are the opposite of taxes: they increase households' disposable income, just as taxes reduce disposable income. Thus, an increase in transfer payments financed by an increase in taxes leaves disposable income unchanged. We can now revise our definition of T to equal taxes minus transfer payments. Disposable income, $Y – T$, includes both the negative impact of taxes and the positive impact of transfer payments.

If government purchases equal taxes minus transfers, then G=T,$G = T$, and the government has a *balanced budget*. If G exceeds T, the government runs a *budget deficit*, which it funds by issuing government debt—that is, by borrowing in the financial markets. If G is less than T, the government runs a *budget surplus*, which it can use to repay some of its outstanding debt.

Here we do not try to explain the political process that leads to a particular fiscal policy—that is, to the level of government purchases and taxes. Instead, we take government purchases and taxes as exogenous variables. To denote that these variables are fixed outside our model of national income, we write

G=G¯.T=T¯.

$$G = \overline{G}\,.$$
$$T = \overline{T}\,.$$

We do, however, want to examine the impact of fiscal policy on the endogenous variables, which are determined within the model. The endogenous variables here are consumption, investment, and the interest rate.

To see how the exogenous variables affect the endogenous variables, we must complete the model. This is the subject of the next section.

3-4 What Brings the Supply and Demand for Goods and Services into Equilibrium?

We have now come full circle in the circular flow diagram, Figure 3-1. We began by examining the supply of goods and services, and we have just discussed the demand for them. How can we be certain that all these flows balance? In other words, what ensures that the sum of consumption, investment, and government purchases equals the amount of output produced? In this classical model, the interest rate is the price that has the crucial role of equilibrating supply and demand.

There are two ways to think about the role of the interest rate in the economy. We can consider how the interest rate affects the supply and demand for goods or services. Or we can consider how the interest rate affects the supply and demand for loanable funds. As we will see, these two approaches are two sides of the same coin.

Equilibrium in the Market for Goods and Services: The Supply and Demand for the Economy’s Output

The following equations summarize the discussion of the demand for goods and services in Section 3-3:

Y=C+I+G.C=C(Y−T).I=I(r).G=G¯.T=T¯.

$$\begin{aligned} Y &= C + I + G. \\ C &= C(Y - T). \\ I &= I(r). \\ G &= \overline{G}. \\ T &= \overline{T}. \end{aligned}$$

The demand for the economy’s output comes from consumption, investment, and government purchases. Consumption depends on disposable income, investment depends on the real interest rate, and government purchases and taxes are the exogenous variables set by fiscal policymakers.

To this analysis, let’s add what we learned about the supply of goods and services in Section 3-1. There we saw that the factors of production and the production function determine the quantity of output supplied to the economy:

$$\begin{aligned} Y &= F(\overline{K}, \overline{L}) \\ &= \overline{Y}. \end{aligned}$$

Now let’s combine these equations describing the supply and demand for output. If we substitute the consumption function and the investment function into the national income accounts identity, we obtain

$$Y = C(Y - T) + I(r) + G.$$

Because the variables *G* and *T* are fixed by policy, and the level of output, *Y*, is fixed by the factors of production and the production function, we can write

$$\overline{Y} = C(\overline{Y} - \overline{T}) + I(r) + \overline{G}.$$

This equation states that the supply of output equals its demand, which is the sum of consumption, investment, and government purchases.

Notice that the interest rate *r* is the only variable not already determined in the last equation. This is because the interest rate still has a key role to play: it must adjust to ensure that the demand for goods equals the supply. The higher the interest rate, the lower the level of investment, and thus the lower the demand for goods and services, $C + I + G$. If the interest rate is too high, then investment is too low, and the demand for output falls short of the supply. If the interest rate is too low, then investment is too high, and the demand exceeds the supply. *At the equilibrium interest rate, the demand for goods and services equals the supply.*

This conclusion may seem mysterious: how does the interest rate get to the level that balances the supply and demand for goods and services? The best way to answer this question is to consider how financial markets fit into the story.

Equilibrium in the Financial Markets: The Supply and Demand for Loanable Funds

Because the interest rate is the cost of borrowing and the return to lending in financial markets, we can better understand the role of the interest rate in the economy by thinking about the financial markets. To do this, rewrite the national income accounts identity as

$$Y - C - G = I.$$

The term $Y - C - G$ is the output that remains after the demands of consumers and the government have been satisfied; it is called **national saving**, or simply **saving** (*S*). In this form, the national income accounts identity shows that saving equals investment.

To understand this identity more fully, we can split national saving into two parts—one part representing the saving of the private sector and the other representing the saving of the government:

$$S = (Y - T - C) + (T - G) = I.$$

The term $(Y - T - C)$ is disposable income minus consumption, which is **private saving**. The term $(T - G)$ is government revenue minus government spending, which is **public saving**. (If government spending exceeds government revenue, then the government runs a budget deficit, and public saving is negative.) National saving is the sum of private and public saving. The circular flow diagram in Figure 3-1 reveals an interpretation of this equation: this equation states that the flows into the financial markets (private and public saving) must balance the flows out of the financial markets (investment).

To see how the interest rate brings financial markets into equilibrium, substitute the consumption function and the investment function into the national income accounts identity:

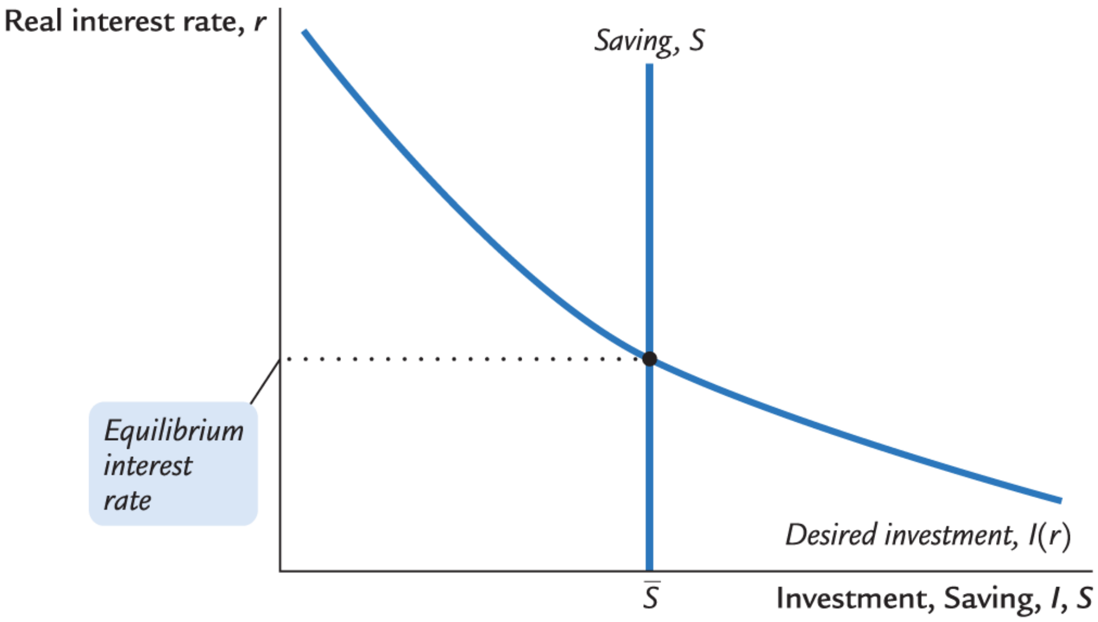
$$Y - C(Y - T) - G = I(r).$$

Next, note that *G* and *T* are fixed by policy, and *Y* is fixed by the factors of production and the production function:

$$\begin{aligned} \overline{Y} - C(\overline{Y} - \overline{T}) - \overline{G} &= I(r) \\ \overline{S} &= I(r). \end{aligned}$$

The left-hand side of this equation shows that national saving depends on income Y and the fiscal-policy variables G and T. For fixed values of Y, G, and T, national saving S is also fixed. The right-hand side of the equation shows that investment depends on the interest rate.

Figure 3-8 graphs saving and investment as a function of the interest rate. The saving function is a vertical line because in this model, saving does not depend on the interest rate. (We relax this assumption later.) The investment function slopes downward: as the interest rate decreases, more investment projects become profitable.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 3-8 Saving, Investment, and the Interest Rate The interest rate adjusts to bring saving and investment into balance. The vertical line represents saving—the supply of loanable funds. The downward-sloping line represents investment—the demand for loanable funds. The intersection of these two curves determines the equilibrium interest rate.

From a quick glance at Figure 3-8, one might think it is a supply-and-demand diagram for a particular good. In fact, saving and investment can be interpreted in terms of supply and demand. In this case, the “good” is **loanable funds**, and its “price” is the interest rate. Saving is the supply of loanable funds: households lend their saving to investors or deposit their saving in a bank that then loans the funds out. Investment is the demand for loanable funds: investors borrow from the public directly by selling bonds or indirectly by borrowing from banks. Because investment depends on the interest rate, the quantity of loanable funds demanded also depends on the interest rate.

The interest rate adjusts until the amount that firms want to invest equals the amount that households want to save. If the interest rate is too low, investors want more of the economy’s output than households want to save. Equivalently, the quantity of loanable funds demanded exceeds the quantity supplied. When this happens, the interest rate rises. Conversely, if the interest rate is too high, households want to save more than firms want to invest; because the quantity of loanable funds supplied is greater than the quantity demanded, the interest rate falls. The equilibrium interest rate is found where the two curves intersect. *At the equilibrium* *interest rate, households' desire to save balances firms' desire to invest, and the quantity of loanable funds supplied equals the quantity demanded.*

Changes in Saving: The Effects of Fiscal Policy

We can use our model to show how fiscal policy affects the economy. When the government changes its spending or the level of taxes, it affects the demand for the economy's output of goods and services and alters national saving, investment, and the interest rate.

An Increase in Government Purchases

Consider first the effects of an increase in government purchases by an amount ΔG. The immediate impact is to increase the demand for goods and services by ΔG. But because total output is fixed by the factors of production, the increase in government purchases must be met by a decrease in some other category of demand. Disposable income $Y - T$ is unchanged, so consumption C is unchanged as well. Therefore, the increase in government purchases must be met by an equal decrease in investment.

To induce investment to fall, the interest rate must rise. Hence, the increase in government purchases causes the interest rate to increase and investment to decrease. Government purchases are said to **crowd out** investment.

To grasp the effects of an increase in government purchases, consider the impact on the market for loanable funds. Because the increase in government purchases is not accompanied by an increase in taxes, the government finances the additional spending by borrowing—that is, by reducing public saving. With private saving unchanged, this government borrowing reduces national saving. As Figure 3-9 shows, a reduction in national saving is represented by a leftward shift in the supply of loanable funds available for investment. At the initial interest rate, the demand for loanable funds exceeds the supply. The equilibrium interest rate rises to the point where the investment schedule crosses the new saving schedule. Thus, an increase in government purchases causes the interest rate to rise from r_1 to r_2.

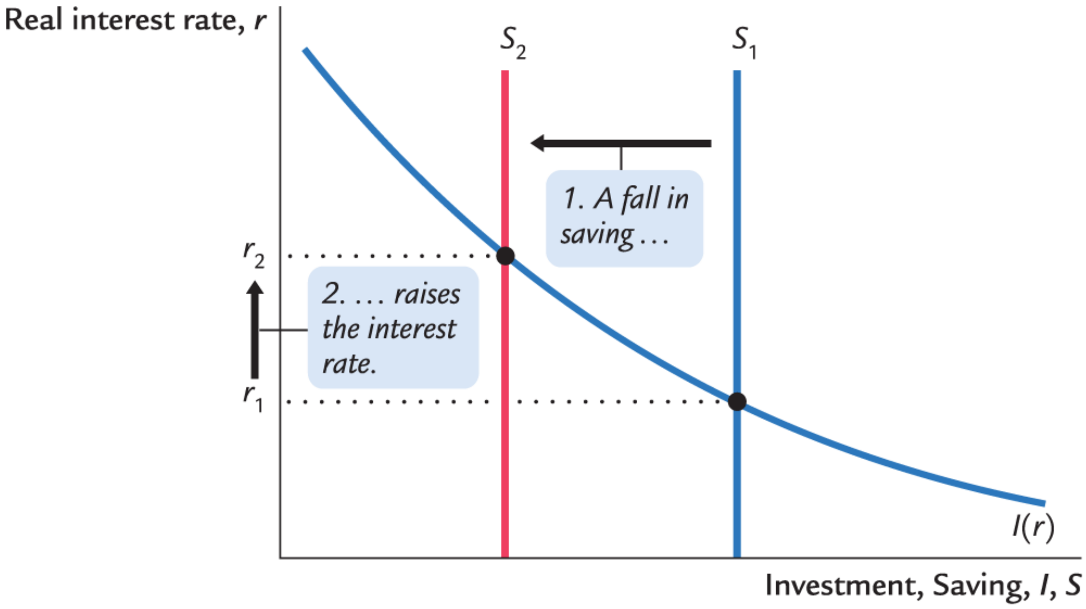


FIGURE 3-9 A Reduction in Saving A reduction in saving, possibly the result of a change in fiscal policy, shifts the saving schedule to the left. The new equilibrium is the point at which the new saving schedule intersects the investment schedule. A reduction in saving lowers the amount of investment and raises the interest rate. Fiscal-policy actions that reduce saving are said to crowd out investment.

A Decrease in Taxes

Now consider a reduction in taxes of ΔT. The immediate impact of the tax cut is to raise disposable income and thus to raise consumption. Disposable income rises by ΔT, and consumption rises by an amount equal to ΔT times the marginal propensity to consume *MPC*. The higher the *MPC*, the greater the impact of the tax cut on consumption.

Because the economy's output is fixed by the factors of production and the level of government purchases is fixed by the government, the increase in consumption must be met by a decrease in investment. For investment to fall, the interest rate must rise. Hence, a reduction in taxes, like an increase in government purchases, crowds out investment and raises the interest rate.

We can also analyze the effect of a tax cut by looking at saving and investment. Because the tax cut raises disposable income by ΔT, consumption goes up by $MPC \times \Delta T$. National saving *S*, which equals $Y - C - G$, falls by the same amount as consumption rises. As in Figure 3-9, the reduction in saving shifts the supply of loanable funds to the left, which increases the equilibrium interest rate and crowds out investment.

Changes in Investment Demand

So far, we have discussed how fiscal policy can change national saving. We can also use our model to examine the other side of the market—the demand for investment. In this section, we look at the causes and effects of changes in investment demand.

One reason investment demand might increase is technological innovation. Suppose, for example, that someone invents a new technology, such as the railroad or the computer. Before a firm or household can take advantage of the innovation, it must buy investment goods. The invention of the railroad had no value until railroad cars were produced and tracks were laid. The idea of the computer was not productive until computers were manufactured. Thus, technological innovation leads to an increase in investment demand.

Investment demand may also change because the government encourages or discourages investment through the tax laws. For example, suppose that the government increases personal income taxes and uses the extra revenue to provide tax cuts for those who invest in new capital. Such a change in the tax laws makes more investment projects profitable and, like a technological innovation, increases the demand for investment goods.

Figure 3-10 shows the effects of an increase in investment demand. At any given interest rate, the demand for investment goods (and also for loanable funds) is higher. This increase in demand is represented by a shift in the investment schedule to the right. The economy moves from the old equilibrium, point A, to the new equilibrium, point B.

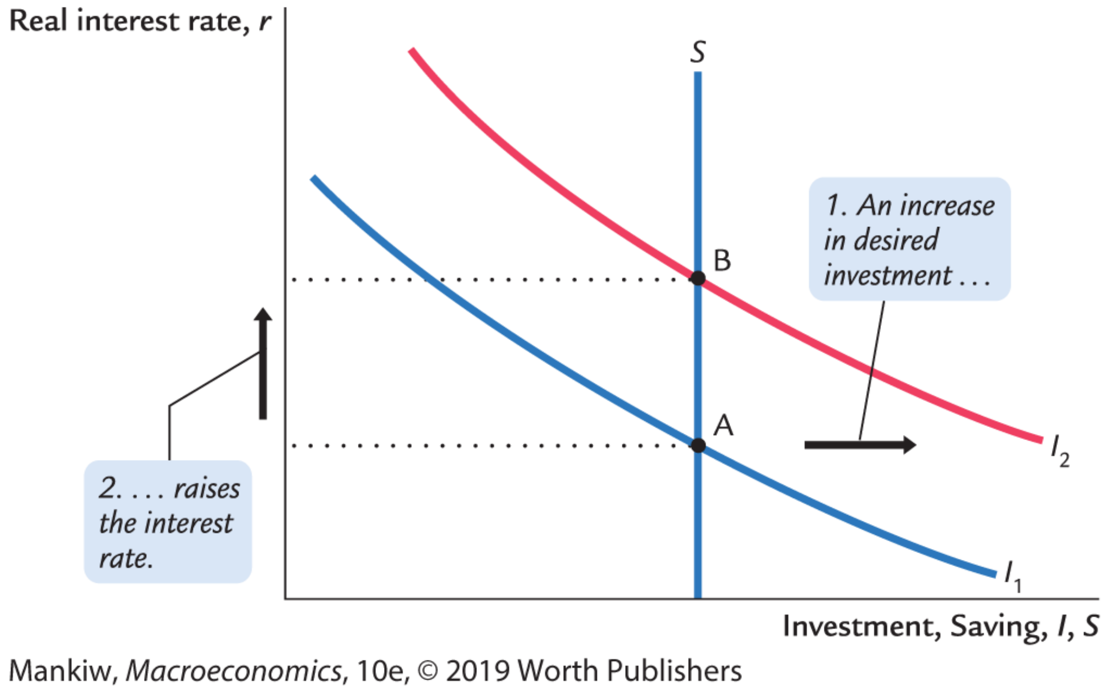


FIGURE 3-10 An Increase in the Demand for Investment An increase in the demand for investment goods shifts the investment schedule to the right. At any given interest rate, the amount of investment is greater. The equilibrium moves from point A to point B. Because the amount of saving is fixed, the increase in investment demand raises the interest rate while leaving the equilibrium amount of investment unchanged.

The surprising implication of Figure 3-10 is that the equilibrium amount of investment is unchanged.

Under our assumptions, the fixed level of saving determines the amount of investment; in other words, there is a fixed supply of loanable funds. An increase in investment demand merely raises the equilibrium interest rate.

We would reach a different conclusion, however, if we modified our simple consumption function and allowed consumption (and its flip side, saving) to depend on the interest rate. Because the interest rate is the return to saving (as well as the cost of borrowing), a higher interest rate might reduce consumption and increase saving. In this case, the saving schedule would be upward sloping rather than vertical.

With an upward-sloping saving schedule, an increase in investment demand would raise both the equilibrium interest rate and the equilibrium quantity of investment. Figure 3-11 shows such a change. The increase in the interest rate causes households to consume less and save more. The decrease in consumption frees resources for investment.

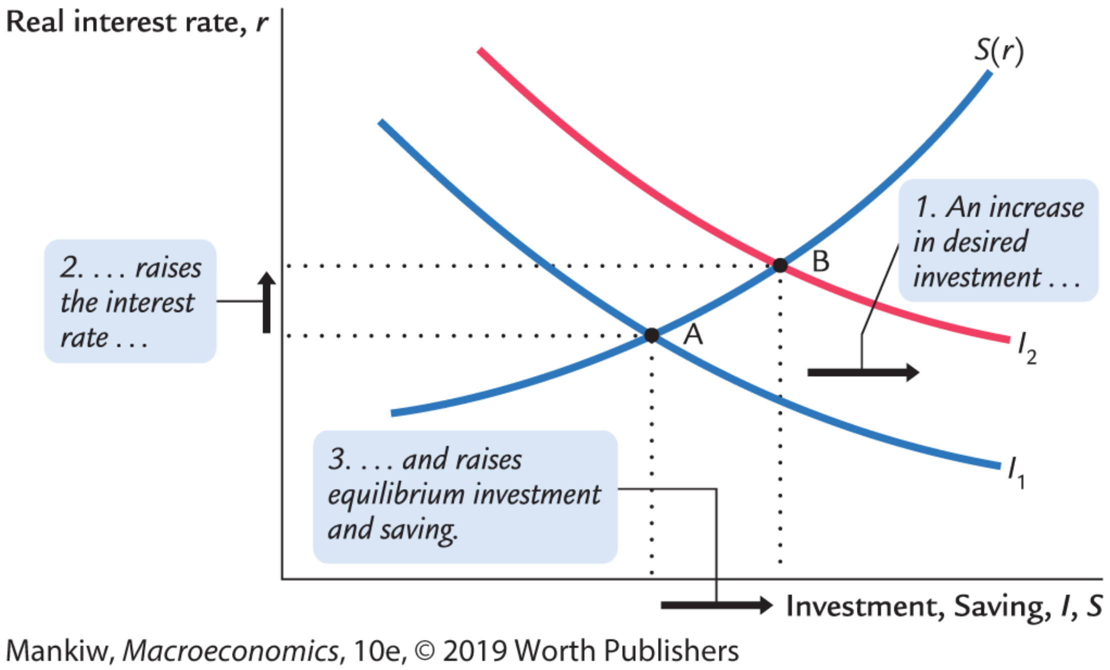


FIGURE 3-11 An Increase in Investment Demand When Saving Depends on the Interest Rate When saving is positively related to the interest rate, a rightward shift in the investment schedule increases the interest rate and the amount of investment. The higher interest rate induces people to increase saving, which in turn allows investment to increase.

3-5 Conclusion

In this chapter, we have developed a model that explains the production, distribution, and allocation of the economy's output of goods and services. The model relies on the classical assumption that prices adjust to equilibrate supply and demand. In this model, factor prices equilibrate factor markets, and the interest rate equilibrates the supply and demand for goods and services (or, equivalently, the supply and demand for loanable funds). Because the model incorporates all the interactions illustrated in the circular flow diagram in Figure 3-1, it is sometimes called a *general equilibrium model*.

Throughout the chapter, we have discussed various applications of the model. The model can explain how income is divided among the factors of production and how factor prices depend on factor supplies. We have also used the model to discuss how fiscal policy alters the allocation of output among its alternative uses—consumption, investment, and government purchases—and how it affects the equilibrium interest rate.

At this point it is useful to review some of the simplifying assumptions we have made, which we will relax in future chapters:

- We have ignored the role of money, the asset with which goods and services are bought and sold. In Chapters 4 and 5, we discuss how money affects the economy and the influence of monetary policy.
- We have assumed that there is no trade with other countries. In Chapter 6, we consider how international interactions affect our conclusions.
- We have assumed that the labor force is fully employed. In Chapter 7, we examine the reasons for unemployment and see how public policy influences the level of unemployment.
- We have assumed that the capital stock, the labor force, and the production technology are fixed. In Chapters 8 and 9, we see how changes over time in each of these lead to growth in the economy's output of goods and services.
- We have ignored the role of short-run sticky prices. In Chapters 10 through 14, we develop a model of short-run fluctuations that includes sticky prices.

Before going on to these chapters, return to the beginning of this one and make sure you can answer the questions about national income that we started with.

CHAPTER 4

The Monetary System: What It Is and How It Works

> There have been three great inventions since the beginning of time: fire, the wheel, and central banking.
>
> —Will Rogers

The two arms of macroeconomic policy are monetary and fiscal policy. Fiscal policy encompasses the government's decisions about spending and taxation, as we saw in the previous chapter. Monetary policy refers to decisions about the nation's system of coin, currency, and banking. Fiscal policy is usually made by elected representatives, such as the U.S. Congress, British Parliament, or Japanese Diet. Monetary policy is made by central banks, which are typically set up by elected representatives but allowed to operate independently. Examples include the U.S. Federal Reserve, the Bank of England, and the Bank of Japan. Will Rogers was exaggerating when he said that central banking was one of the three greatest inventions of all time, but he was right in suggesting that these policymaking institutions have a major influence over the lives and livelihoods of people around the world.

Much of this book is aimed at explaining the effects and proper role of monetary and fiscal policy. This chapter begins our analysis of monetary policy. We address three related questions. First, what is money? Second, what is the role of a nation's banking system in determining the amount of money in the economy? Third, how does a nation's central bank influence the banking system and the money supply?

This chapter's introduction to the monetary system provides the foundation for understanding monetary policy. In the next chapter, consistent with the long-run focus of this part of the book, we examine the long-run effects of monetary policy. The short-run effects of monetary policy are more complex. We start discussing that topic in Chapter 10, but it will take several chapters to develop a complete explanation. This chapter gets us ready. Both the long-run and short-run analysis of monetary policy must be grounded in an understanding of what money is, how banks affect it, and how central banks control it.

4-1 What Is Money?

When we say that a person has a lot of money, we usually mean that he is wealthy. By contrast, economists use the term "money" in a more specialized way. To an economist, money does not refer to all wealth but only to one type of it: **money** is the stock of assets that can be readily used to make transactions. Roughly speaking, the dollars (or, in other countries, pesos, pounds, or yen) in the hands of the public make up the nation's stock of money.

The Functions of Money

Money has three purposes: it is a store of value, a unit of account, and a medium of exchange.

As a **store of value**, money is a way to transfer purchasing power from the present to the future. If you work today and earn $100, you can hold the money and spend it tomorrow, next week, or next month. Money is not a perfect store of value: if prices are rising, the amount you can buy with any given quantity of money is falling. Even so, people hold money because they can trade it for goods and services at some time in the future.

As a **unit of account**, money provides the terms in which people quote prices and record debts. Microeconomics teaches that resources are allocated according to relative prices—the prices of goods relative to other goods—yet stores post their prices in dollars and cents. A car dealer says that a car costs $40,000, not 800 shirts (even though it may amount to the same thing). Similarly, most debts require the debtor to deliver a certain number of dollars in the future, not an amount of some commodity. Money is the yardstick with which we measure economic transactions.

As a **medium of exchange**, money is what people use to buy goods and services. "This note is legal tender for all debts, public and private" is printed on the U.S. dollar. When you walk into stores, you are confident that the shopkeepers will accept your money in exchange for the items they are selling. The ease with which an asset can be converted into the medium of exchange and used to buy other things (goods, services, or capital assets) is called the asset's *liquidity*. Because money is the medium of exchange, it is the economy's most liquid asset.

To better understand the functions of money, try to imagine an economy without it: a barter economy. In such a world, trade requires the *double coincidence of wants*—the unlikely happenstance of two people each having a good that the other wants at the right time and place to make an exchange. A barter economy permits only simple transactions.

Money makes more complex transactions possible. A professor uses his salary to buy books; the book publisher uses its revenue from the sale of books to buy paper; the paper company uses its revenue from the sale of paper to buy wood that it grinds into paper pulp; the lumber company uses revenue from the sale of wood to pay the lumberjack; the lumberjack uses his income to send his child to college; and the college uses its tuition receipts to pay the salary of the professor. In a modern economy, trade often involves many parties and is facilitated by the use of money.

The Types of Money

Money takes many forms. In the U.S. economy, we make transactions with an item whose sole function is to act as money: dollar bills. These pieces of green paper with small portraits of famous Americans would have little value if they were not widely accepted as money. Money without intrinsic value is called **fiat money** because it is established as money by government decree, or fiat.

"And how would you like your funny money?"
Bernard Schoenbaum/The New Yorker/Conde Nast/The Cartoon Bank

Fiat money is the norm in most economies today, but many societies in the past have used a commodity with some intrinsic value for money. This type of money is called **commodity money**. The most widespread example is gold. When people use gold as money (or use paper money redeemable for gold), the economy is said to be on a **gold standard**. Gold is a form of commodity money because it can be used for various purposes—jewelry, dental fillings, and so on—as well as for transactions. The gold standard was common throughout the world during the late nineteenth century.

CASE STUDY

Money in a POW Camp

An unusual form of commodity money developed in some Nazi prisoner of war (POW) camps during World War

II. The Red Cross supplied the prisoners with various goods—food, clothing, cigarettes, and so on. Yet these rations were allocated without close attention to personal preferences, so the allocations were often inefficient. One prisoner might have preferred chocolate, while another might have preferred cheese, and a third might have wanted a new shirt. The differing tastes and endowments of the prisoners led them to trade with one another.

Barter was an inconvenient way to allocate these resources, however, because it required the double coincidence of wants. In other words, a barter system was not the easiest way to ensure that each prisoner received the goods he valued most. Even the limited economy of the POW camp needed money to facilitate exchange.

Eventually, cigarettes became the established "currency" in which prices were quoted and with which trades were made. A shirt, for example, cost about 80 cigarettes. Services were also quoted in cigarettes: some prisoners offered to do other prisoners' laundry for two cigarettes per garment. Even nonsmokers were happy to accept cigarettes in exchange, knowing they could trade the cigarettes in the future for some good they did enjoy. Within the POW camp the cigarette became the store of value, the unit of account, and the medium of exchange.[1] ■

The Development of Fiat Money

It is not surprising that in any society, no matter how primitive, some form of commodity money arises to facilitate exchange: people are willing to accept a commodity currency such as gold because it has intrinsic value. Fiat money, however, is more perplexing. What would make people start valuing something that is intrinsically useless?

To understand how the evolution from commodity money to fiat money takes place, imagine an economy in which people carry around bags of gold. When making a purchase, the buyer measures out the appropriate amount of gold. If the seller is convinced that the weight and purity of the gold are right, the exchange is made.

The government might first get involved in the monetary system to help people reduce transaction costs. Using raw gold as money is costly because it takes time to verify the purity of the gold and to measure the correct quantity. To reduce these costs, the government can mint gold coins of known purity and weight. The coins are more convenient than gold bullion because their values are widely recognized.

The next step is for the government to accept gold from the public in exchange for gold certificates—pieces of paper that can be redeemed for a certain quantity of gold. If people believe the government's promise to redeem the paper bills for gold, the bills are just as valuable as the gold itself. In addition, because the bills are lighter than gold (and gold coins), they are easier to use in transactions. Eventually, no one carries gold around at all, and these gold-backed government bills become the monetary standard.

Finally, the gold backing becomes irrelevant. If no one ever bothers to redeem the bills for gold, no one cares if the option is abandoned. As long as everyone accepts the paper bills in exchange, they will have value and serve as money. Thus, the system of commodity money evolves into a system of fiat money. In the end, the use of money in exchange is a social convention: everyone values fiat money because they expect everyone else to value it.

CASE STUDY

Money and Social Conventions on the Island of Yap

The economy of Yap, a small island in the Pacific, once had a type of money that was something between commodity and fiat money. The traditional medium of exchange in Yap was *fei*, stone wheels up to 12 feet in diameter. These stones had holes in the center so that they could be carried on poles and used for exchange.

Large stone wheels are not a convenient form of money. The stones were heavy, so it took substantial effort for a new owner to take his *fei* home after completing a transaction. Although the monetary system facilitated exchange, it did so at great cost.

Eventually, it became common practice for the new owner of the *fei* not to bother to take physical possession of the stone. Instead, the new owner accepted a claim to the *fei* without moving it. In future bargains, he traded this claim for goods that he wanted. Having physical possession of the stone became less important than having legal claim to it.

This practice was put to a test when a valuable stone was lost at sea during a storm. Because the owner lost his money by accident rather than through negligence, everyone agreed that his claim to the *fei* remained valid. Generations later, when no one alive had ever seen this stone, the claim to this *fei* was still valued in exchange.

Even today, stone money is still valued on the island. But it is not the medium of exchange used for most routine transactions. For that purpose, the 11,000 residents of Yap use something more prosaic: the U.S. dollar.[2] ■

FYI

Bitcoin: The Strange Case of a Digital Money

In 2009, the world was introduced to a new and unusual asset, called *bitcoin*. Conceived by an anonymous computer expert (or group of experts) who goes by the name Satoshi Nakamoto, bitcoin is intended to be a form of money that exists only in electronic form. Individuals originally obtain bitcoins by using computers to solve complex mathematical problems. The bitcoin protocol is designed to limit the number of bitcoins that can ever be “mined” in this way to 21 million units (though experts disagree whether the number of bitcoins is truly limited). After the bitcoins are created, they can be used in exchange. They can be bought and sold for U.S. dollars and other currencies on organized bitcoin exchanges, where the exchange rate is set by supply and demand. You can use bitcoins to buy things from any vendor who is willing to accept them.

As a form of money, bitcoins are neither commodity money nor fiat money. Unlike commodity money, they have no intrinsic value. You can’t use bitcoins for anything other than exchange. Unlike fiat money, they are not created by government decree. Indeed, many fans of bitcoin embrace the fact that this electronic cash exists apart from government. (Some users of it are engaged in illicit transactions such as the drug trade and, therefore, like the anonymity that bitcoin transactions offer.) Bitcoins have value only to the extent that people accept the social convention of taking them in exchange. From this perspective, the modern bitcoin resembles the primitive money of Yap.

Throughout its brief history, the value of a bitcoin, as measured by its price in U.S. dollars, has fluctuated wildly. Throughout 2010, the price of a bitcoin ranged from 5 cents to 39 cents. In 2011 the price rose to above $1, and in 2013 it briefly rose above $1,000 before falling below $500 in 2014. Over the next few years, it skyrocketed, reaching more than $15,000 in 2017. Gold is often considered a risky asset, but the day-to-day volatility of bitcoin prices has been several times the volatility of gold prices.

The long-term success of bitcoin depends on whether it succeeds in performing the functions of money: a store of value, a unit of account, and a medium of exchange. Many economists are skeptical that it will do these tasks well. Bitcoin's volatility makes it a risky way to hold wealth and an inconvenient measure in which to post prices. At least so far, few retailers accept it in exchange, and those that do have only a small volume of bitcoin transactions.

Advocates of bitcoin see it as the money of the future. Another possibility, however, is that it is a speculative fad that will eventually run its course.[3]

How the Quantity of Money Is Controlled

The quantity of money available in an economy is called the **money supply**. In a system of commodity money, the money supply is simply the quantity of that commodity. In an economy that uses fiat money, such as most economies today, the government controls the supply of money: legal restrictions give the government a monopoly on the printing of money. Just as the levels of taxation and government purchases are policy instruments of the government, so is the quantity of money. The government's control over the money supply is called **monetary policy**.

In most countries, monetary policy is delegated to a partially independent institution called the **central bank**. The central bank of the United States is the **Federal Reserve**—often called *the Fed*. If you look at a U.S. dollar bill, you will see that it is called a *Federal Reserve Note*. Decisions about monetary policy are made by the Fed's Federal Open Market Committee (FOMC). This committee consists of two groups: (1) members of the Federal Reserve Board, who are appointed by the president and confirmed by the Senate, and (2) the presidents of the regional Federal Reserve Banks, who are chosen by these banks' boards of directors. The FOMC meets about every six weeks to discuss and set monetary policy.

The main way in which the Fed controls the supply of money is through **open-market operations**—the purchase and sale of government bonds. When the Fed wants to increase the money supply, it uses some of the dollars it has to buy government bonds from the public. Because these dollars leave the Fed and enter the hands of the public, the purchase increases the quantity of money in circulation. Conversely, when the Fed wants to decrease the money supply, it sells some government bonds from its own portfolio. This open-market sale of bonds takes some dollars out of the hands of the public and, thus, decreases the quantity of money in circulation. (Later in the chapter, we explore in more detail how the Fed controls the supply of money.)

How the Quantity of Money Is Measured

One of our goals is to determine how the money supply affects the economy; we turn to that topic in the next chapter. As a background for that analysis, let’s first discuss how economists measure the quantity of money.

Because money is the stock of assets used for transactions, the quantity of money is the quantity of those assets. In simple economies, this quantity is easy to measure. In the POW camp, the quantity of money was the number of cigarettes in the camp. On the island of Yap, the quantity of money was the number of *fei* on the island. But how can we measure the quantity of money in more complex economies? The answer is not obvious, because no single asset is used for all transactions. People can transact using various assets, such as cash in their wallets or deposits in their checking accounts, although some assets are more convenient to use than others.

The most obvious asset to include in the quantity of money is **currency**, the sum of outstanding paper money and coins. Many day-to-day transactions use currency as the medium of exchange.

A second type of asset used for transactions is **demand deposits**, the funds people hold in their checking accounts. If most sellers accept personal checks or debit cards that access checking accounts balances, then assets in these accounts are almost as convenient as currency. That is, the assets are in a form that can easily facilitate a transaction. Demand deposits are therefore added to currency when measuring the quantity of money.

Once we admit the logic of including demand deposits in the measured money stock, many other assets become candidates for inclusion. Funds in savings accounts, for example, can be easily transferred into checking accounts or accessed by debit cards; these assets are almost as convenient for transactions. Money market mutual funds allow investors to write checks against their accounts, although restrictions sometimes apply regarding the size of the check or number of checks written. Because these assets can be easily used for transactions, they should arguably be included in the quantity of money.

Because it is hard to judge which assets should be included in the money stock, more than one measure is available. Table 4-1 presents the three measures of the money stock that the Federal Reserve calculates for the U.S. economy, along with a list of assets included in each measure. From the smallest to the largest, they are denoted *C*, *M*1, and *M*2. The most common measures for studying the effects of money on the economy are *M*1 and *M*2.

TABLE 4-1 The Measures of Money

Symbol	Assets Included	Amount in July 2017 (billions of dollars)
C	Currency	$ 1,486
*M*1	Currency plus demand deposits, traveler's checks, and other checkable deposits	3,528
*M*2	*M*1 plus retail money market mutual fund balances, saving deposits (including money market deposit accounts), and small time deposits	13,602

Data from: Federal Reserve.

FYI

How Do Credit Cards and Debit Cards Fit into the Monetary System?

Many people use credit or debit cards to make purchases. Because money is the medium of exchange, one might naturally wonder how these cards fit into the measurement and analysis of money.

Let's start with credit cards. One might guess that credit cards are part of the economy's stock of money. In fact, however, measures of the money stock do not take credit cards into account because credit cards are not really a method of payment but a method of *deferring* payment. When you buy an item with a credit card, the bank that issued the card pays the store what it is due. Later, you repay the bank. When the time comes to pay your credit card bill, you will likely do so by transferring funds from your checking account, either electronically or by writing a check. The balance in this checking account is part of the economy's stock of money.

The story is different with debit cards, which automatically withdraw funds from a bank account to pay for items bought. Rather than allowing users to postpone payment for their purchases, a debit card gives users immediate access to deposits in their bank accounts. Using a debit card is like writing a check. The account balances that lie behind debit cards are included in measures of the quantity of money.

Even though credit cards are not a form of money, they are still important for analyzing the monetary system. Because people with credit cards can pay many of their bills all at once at the end of the month, rather than sporadically as they make purchases, they may hold less money on average than people without credit cards. Thus, the increased popularity of credit cards may reduce the amount of money that people choose to hold. In other words, credit cards are not part of the supply of money, but they may affect the demand for money.

4-2 The Role of Banks in the Monetary System

Earlier, we introduced the concept of “money supply” in a highly simplified manner. We defined the quantity of money as the number of dollars held by the public, and we assumed that the Federal Reserve controls the money supply by changing the number of dollars in circulation through open-market operations. This explanation was a good starting point for understanding what determines the supply of money, but it is incomplete because it omits the role of the banking system in this process.

In this section, we see that the money supply is determined not only by Fed policy but also by the behavior of households (which hold money) and banks (in which money is held). We begin by recalling that the money supply includes both currency in the hands of the public and deposits (such as checking account balances) at banks that households can use on demand for transactions. If *M* denotes the money supply, *C* currency, and *D* demand deposits, we can write

Money Supply=Currency+Demand DepositsM=C+D

$$\begin{array}{ccccc} \text{Money Supply} & = & \text{Currency} & + & \text{Demand Deposits} \\ M & = & C & + & D \end{array}$$

To understand the money supply, we must understand the interaction between currency and demand deposits and how the banking system, together with Fed policy, influences these two components of the money supply.

100-Percent-Reserve Banking

We begin by imagining a world without banks. In such a world, all money takes the form of currency, and the quantity of money is simply the amount of currency that the public holds. For this discussion, suppose that there is $1,000 of currency in the economy.

Now introduce banks. At first, suppose that banks accept deposits but do not make loans. The only purpose of the banks is to provide a safe place for depositors to keep their money.

The deposits that banks have received but have not lent out are called **reserves**. Some reserves are held in the vaults of local banks throughout the country, but most are held at a central bank, such as the Federal Reserve. In our hypothetical economy, all deposits are held as reserves: banks simply accept deposits, place the money in reserve, and leave the money there until the depositor makes a withdrawal or writes a check against the balance. This system is called **100-percent-reserve banking**.

Suppose that households deposit the economy's entire $1,000 in Firstbank. Firstbank's **balance sheet**—its accounting statement of assets and liabilities—looks like this:

Firstbank's Balance Sheet

Assets		Liabilities	
Reserves	$1,000	Deposits	$1,000

The bank's assets are the $1,000 it holds as reserves; the bank's liabilities are the $1,000 it owes to depositors. Unlike banks in our economy, this bank is not making loans, so it will not earn profit from its assets. The bank presumably charges depositors a small fee to cover its costs.

What is the money supply in this economy? Before the creation of Firstbank, the money supply was the $1,000 of currency. After the creation of Firstbank, the money supply is the $1,000 of demand deposits. A dollar deposited in a bank reduces currency by one dollar and raises deposits by one dollar, so the money supply remains the same. *If banks hold 100 percent of deposits in reserve, the banking system does not affect the supply of money.*

Fractional-Reserve Banking

Now imagine that banks start lending out some of their deposits—for example, to families buying houses or to firms investing in new plants and equipment. The advantage to banks is that they can charge interest on the loans. The banks must keep some reserves on hand so that reserves are available whenever depositors want to make withdrawals. But as long as the amount of new deposits approximately equals the amount of withdrawals, a bank need not keep all its deposits in reserve. Thus, bankers have an incentive to lend. When they do so, we have **fractional-reserve banking**, a system under which banks keep only a fraction of their deposits in reserve.

Here is Firstbank's balance sheet after it makes a loan:

Firstbank's Balance Sheet

Assets		Liabilities	
Reserves	$200	Deposits	$1,000
Loans	$800		

This balance sheet assumes that the *reserve–deposit ratio*—the fraction of deposits kept in reserve—is 20 percent. Firstbank keeps $200 of the $1,000 in deposits in reserve and lends out the remaining $800.

Notice that Firstbank increases the supply of money by $800 when it makes this loan. Before the loan is made, the money supply is $1,000, equaling the deposits in Firstbank. After the loan is made, the money supply is $1,800: the depositor still has a demand deposit of $1,000, but now the borrower holds $800 in currency. *Thus, in a system of fractional-reserve banking, banks create money*.

The creation of money does not stop with Firstbank. If the borrower deposits the $800 in another bank (or if the borrower uses the $800 to pay someone who then deposits it), the process of money creation continues. Here is the balance sheet of Secondbank:

Secondbank's Balance Sheet

Assets		Liabilities	
Reserves	$160	Deposits	$800
Loans	$640		

Secondbank receives the $800 in deposits, keeps 20 percent, or $160, in reserve and then lends $640. Thus, Secondbank creates $640 of money. If this $640 is eventually deposited in Thirdbank, this bank keeps 20 percent, or $128, in reserve and lends $512, resulting in this balance sheet:

Thirdbank's Balance Sheet

Assets		Liabilities	
Reserves	$128	Deposits	$640
Loans	$512		

The process goes on and on. With each deposit and loan, more money is created.

This process of money creation can continue forever, but it does not create an infinite amount of money. Letting *rr* denote the reserve–deposit ratio, the amount of money that the original $1,000 creates is

$$\text{Original Deposit} = \$1{,}000 \quad \text{Firstbank Lending} = (1 - rr) \times \$1{,}000 \quad \text{Secondbank Lending} = (1 - rr)^2 \times \$1{,}000 \quad \text{Thirdbank}$$

$$[\, 1 + (1 - rr) + (1 - rr)^2 + (1 - rr)^3 + \ldots \,] \times \$1{,}000 = (1/rr) \times \$1{,}000.$$

$$\begin{aligned}
\text{Original Deposit} &= \$1{,}000 \\
\text{Firstbank Lending} &= (1 - rr) \times \$1{,}000 \\
\text{Secondbank Lending} &= (1 - rr)^2 \times \$1{,}000 \\
\text{Thirdbank Lending} &= (1 - rr)^3 \times \$1{,}000 \\
\hline
\text{Total Money Supply} &= \left[1 + (1 - rr) + (1 - rr)^2 + (1 - rr)^3 + \ldots\right] \times \$1{,}000 \\
&= (1/rr) \times \$1{,}000.
\end{aligned}$$

Each \$1 of reserves generates \$(1/*rr*) of money. In our example, $rr = 0.2$, so the original \$1,000 generates \$5,000 of money.[4]

The banking system's ability to create money is the main difference between banks and other financial institutions. As we first discussed in Chapter 3, financial markets have the important function of transferring the economy's resources from those households that wish to save some of their income for the future to those households and firms that wish to borrow to buy investment goods to be used in future production. The process of transferring funds from savers to borrowers is called **financial intermediation**. Many institutions act as financial intermediaries: the most prominent examples are the stock market, the bond market, and the banking system. Yet, of these financial institutions, only banks have the legal authority to create assets (such as checking accounts) that are part of the money supply. Therefore, banks are the only financial institutions that directly influence the money supply.

Note that although the system of fractional-reserve banking creates money, it does not create wealth. When a bank lends some of its reserves, it gives borrowers the ability to make transactions and therefore increases the money supply. The borrowers are also undertaking debt obligations to the bank, however, so the loans do not make them wealthier. In other words, the creation of money by the banking system increases the economy's liquidity, not its wealth.

Bank Capital, Leverage, and Capital Requirements

The model of the banking system presented so far is simplified. That is not necessarily a problem; after all, all models are simplified. But one particular simplifying assumption is noteworthy.

In the bank balance sheets we just examined, a bank takes in deposits and either uses them to make loans or holds them as reserves. Based on this discussion, you might think that it does not take any resources to open a bank. That is, however, not true. Opening a bank requires some capital. That is, the bank owners must start with some financial resources to get the business going. Those resources are called **bank capital** or, equivalently, the equity of the bank's owners.

Here is what a more realistic balance sheet for a bank would look like:

Realbank's Balance Sheet

Assets		Liabilities and Owners' Equity	
Reserves	$200	Deposits	$750
Loans	$500	Debt	$200
Securities	$300	Capital (owners' equity)	$50

The bank obtains resources from its owners who provide capital, from customers by taking in deposits, and from investors by issuing debt. It uses these resources in three ways. Some funds are held as reserves; some are used to make bank loans; and some are used to buy financial securities, such as government or corporate bonds. The bank allocates its resources among these asset classes, considering the risk and return that each offers and any regulations that restrict its choices. The reserves, loans, and securities on the left side of the balance sheet must equal, in total, the deposits, debt, and capital on the right side of the balance sheet.

This business strategy relies on a phenomenon called **leverage**, which is the use of borrowed money to supplement existing funds for purposes of investment. The *leverage ratio* is the ratio of the bank's total assets (the sum of the left side of the balance sheet) to the bank's capital (the one item on the right side of the balance sheet that represents the owners' equity). In this example, the leverage ratio is $1000/$50, or 20. This means that for every dollar of capital that the bank owners have contributed, the bank has $20 of assets and, thus, $19 of deposits and debts.

Because of leverage, a bank can lose capital quickly in tough times. To see how, let's continue with this example. If the bank's assets fall in value by just 5 percent, then the $1,000 of assets is now worth only $950. Since the depositors and debt holders have the legal right to be paid first, the owners' equity falls to zero. That is, when the leverage ratio is 20, a 5 percent fall in the value of the bank assets causes a 100 percent fall in bank capital. If the value of the assets declines by more than 5 percent, assets fall below liabilities, sending bank capital below zero. The bank is said to be *insolvent*. The fear that bank capital may run out, and thus that depositors might not be repaid in full, is what generates bank runs when there is no deposit insurance.

Bank regulators require that banks hold sufficient capital. The goal of a **capital requirement** is to ensure that banks will be able to pay off their depositors and other creditors. The amount of capital required depends on the kind of assets a bank holds. If the bank holds safe assets such as government bonds, regulators require less capital than if the bank holds risky assets such as loans to borrowers whose credit is of dubious quality.

The arcane issues of bank capital and leverage are usually left to bankers, regulators, and financial experts, but they became prominent topics of public debate during and after the financial crisis of 2008–2009. During this period, declining house prices caused many banks and other financial institutions to incur losses on mortgage-backed securities. Because of leverage, the losses to bank capital were proportionately much larger than the losses to bank assets. Some institutions became insolvent. These events had repercussions not only within the financial system but throughout the economy. In the aftermath of the crisis, many observers suggested that banks be subject to higher capital requirements.[5]

For now, we can put aside the issues of bank capital and leverage. But they will resurface when we discuss financial crises in Chapters 12 and 18.

4-3 How Central Banks Influence the Money Supply

Having seen what money is and how the banking system affects the amount of money in the economy, we are ready to examine how the central bank influences the banking system and the money supply. This influence is the essence of monetary policy.

A Model of the Money Supply

If the Federal Reserve adds a dollar to the economy and that dollar is held as currency, the money supply increases by exactly one dollar. But as we have seen, if that dollar is deposited in a bank, and banks hold only a fraction of their deposits in reserve, the money supply increases by more than one dollar. As a result, to understand what determines the money supply under fractional-reserve banking, we need to take account of the interactions among (1) the Fed's decision about how many dollars to create, (2) banks' decisions about whether to hold deposits as reserves or to lend them out, and (3) households' decisions about whether to hold their money in the form of currency or demand deposits. This section develops a model of the money supply that includes all these factors.

The model has three exogenous variables:

- The **monetary base** *B* is the total number of dollars held by the public as currency *C* and by the banks as reserves *R*. It is directly controlled by the Federal Reserve.
- The **reserve–deposit ratio** *rr* is the fraction of deposits that banks hold in reserve. It is determined by the business policies of banks and the laws regulating banks.
- The **currency–deposit ratio** *cr* is the amount of currency *C* people hold as a fraction of their holdings of demand deposits *D*. It reflects the preferences of households about the form of money they wish to hold.

By showing how the money supply depends on the monetary base, the reserve–deposit ratio, and the currency–deposit ratio, this model is useful for understanding how Fed policy and the choices of banks and households influence the money supply.

We begin with the definitions of the money supply and the monetary base:

M=C+D,B=C+R.

$$M = C + D,$$
$$B = C + R.$$

The first equation states that the money supply is the sum of currency and demand deposits. The second equation states that the monetary base is the sum of currency and bank reserves. To solve for the money supply as a function of the three exogenous variables (*B*, *rr*, and *cr*), we divide the first equation by the second to obtain

$$\frac{M}{B} = \frac{C+D}{C+R}.$$

We then divide both the top and bottom of the expression on the right by *D*.

$$\frac{M}{B} = \frac{C/D+1}{C/D+R/D}.$$

Note that *C*/*D* is the currency–deposit ratio *cr* and that *R*/*D* is the reserve–deposit ratio *rr*. Making these substitutions, and bringing the *B* from the left to the right side of the equation, we obtain

$$M = \frac{cr+1}{cr+rr} \times B.$$

This equation shows how the money supply depends on the three exogenous variables.

We can now see that the money supply is proportional to the monetary base. The factor of proportionality, (*cr* + 1)/(*cr* + *rr*), is denoted *m* and is called the **money multiplier**. We can write

$$M = m \times B.$$

Each dollar of the monetary base produces *m* dollars of money. Because the monetary base has a multiplied effect on the money supply, the monetary base is sometimes called **high-powered money**.

Here's a numerical example. Suppose that the monetary base *B* is $800 billion, the reserve–deposit ratio *rr* is 0.1, and the currency–deposit ratio *cr* is 0.8. In this case, the money multiplier is

$$m = \frac{0.8+1}{0.8+0.1} = 2.0,$$

and the money supply is

$$M=2.0\times\$800\text{ billion}=\$1{,}600\text{ billion}.$$

Each dollar of the monetary base generates two dollars of money, so the total money supply is $1,600 billion.

We can now see how changes in the three exogenous variables—*B*, *rr*, and *cr*—cause the money supply to change.

1. The money supply is proportional to the monetary base. Thus, an increase in the monetary base increases the money supply by the same percentage.
2. The lower the reserve–deposit ratio, the more loans banks make, and the more money banks create from every dollar of reserves. Thus, a decrease in the reserve–deposit ratio raises the money multiplier and the money supply.
3. The lower the currency–deposit ratio, the fewer dollars of the monetary base the public holds as currency, the more base dollars banks hold as reserves, and the more money banks can create. Thus, a decrease in the currency–deposit ratio raises the money multiplier and the money supply.

With this model in mind, we can discuss the ways in which the Fed influences the money supply.

The Instruments of Monetary Policy

Although it is often convenient to make the simplifying assumption that the Federal Reserve controls the money supply directly, in fact the Fed controls the money supply indirectly using various instruments. These instruments can be classified into two broad groups: those that influence the monetary base and those that influence the reserve–deposit ratio and thereby the money multiplier.

How the Fed Changes the Monetary Base

As we discussed earlier, *open-market operations* are the purchases and sales of government bonds by the Fed. When the Fed buys bonds from the public, the dollars it pays for the bonds increase the monetary base and thereby increase the money supply. When the Fed sells bonds to the public, the dollars it receives reduce the monetary base and thus decrease the money supply. Open-market operations are the policy instrument that the Fed uses most often. In fact, the Fed conducts open-market operations in New York bond markets almost every weekday.

The Fed can also alter the monetary base and the money supply by lending reserves to banks. Banks borrow from the Fed when they think they do not have enough reserves on hand, either to satisfy bank regulators, meet depositor withdrawals, make new loans, or satisfy some other business requirement. When the Fed lends to a bank that is having trouble obtaining funds from elsewhere, it is said to act as the *lender of last resort*.

Banks can borrow from the Fed in various ways. Traditionally, banks have borrowed at the Fed's so-called *discount window*; the **discount rate** is the interest rate that the Fed charges on these loans. The lower the discount rate, the cheaper are borrowed reserves, and the more banks borrow at the Fed's discount window. Hence, a reduction in the discount rate raises the monetary base and the money supply.

In response to the financial crisis of 2008–2009, the Federal Reserve set up several new mechanisms for banks to borrow from it. For example, under the *Term Auction Facility*, the Fed set a quantity of funds it wanted to lend to banks, and eligible banks then bid to borrow those funds. The loans went to the highest eligible bidders—that is, to the banks that had acceptable collateral and offered to pay the highest interest rate. Unlike at the discount window, where the Fed sets the price of a loan and the banks determine the quantity of borrowing, at the Term Auction Facility the Fed set the quantity of borrowing and a competitive bidding process among banks determined the price. The last Term Auction Facility auction was conducted in 2010, but this policy illustrates that the Federal Reserve has various ways to alter the monetary base and the money supply.

How the Fed Changes the Reserve–Deposit Ratio

As our model of the money supply shows, the money multiplier is the link between the monetary base and the money supply. The money multiplier depends on the reserve–deposit ratio, which in turn is influenced by various Fed policy instruments.

Reserve requirements are Fed regulations that impose a minimum reserve–deposit ratio on banks. An increase in reserve requirements tends to raise the reserve–deposit ratio and thus lower the money multiplier and the money supply. Changes in reserve requirements are the least frequently used of the Fed's policy instruments. Moreover, in recent years, this tool has become less effective because many banks hold more reserves than are required. Reserves above the minimum required are called **excess reserves**.

In October 2008, the Fed started paying **interest on reserves**. That is, when a bank holds reserves on deposit at the Fed, the Fed now pays the bank interest on those deposits. This change gives the Fed another tool with which to influence the economy. The higher the interest rate on reserves, the more reserves banks will choose to hold. Thus, an increase in the interest rate on reserves will tend to increase the reserve–deposit ratio, lower the money multiplier, and lower the money supply.

CASE STUDY

Quantitative Easing and the Exploding Monetary Base

Figure 4-1 shows the monetary base from 1960 to 2017. You can see that something extraordinary happened after 2007. From 1960 to 2007, the monetary base grew gradually over time. But then from 2007 to 2014 it spiked up substantially, increasing about 5-fold over just a few years.

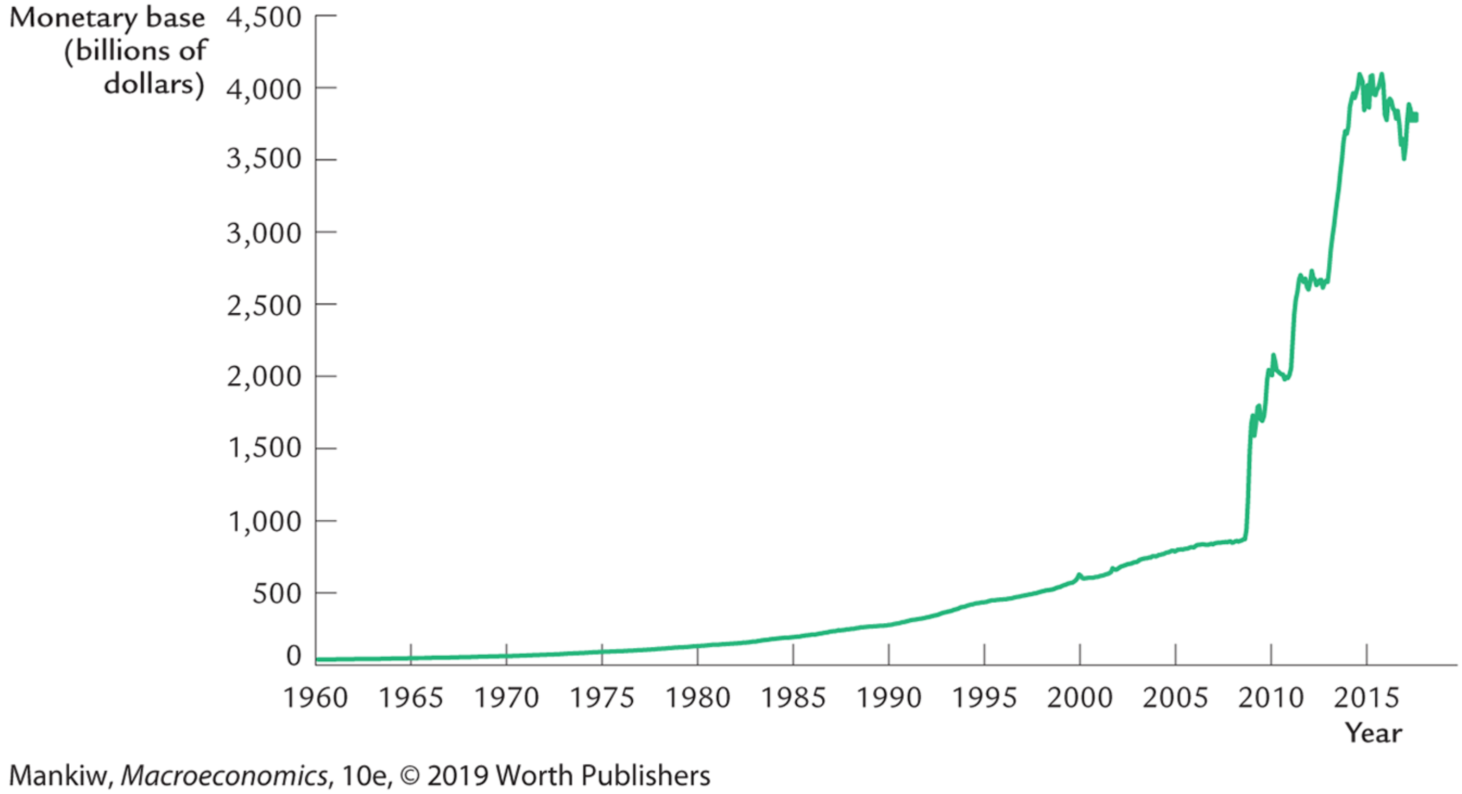


FIGURE 4-1 The Monetary Base The monetary base has historically grown relatively smoothly over time, but from 2007 to 2014 it increased approximately 5-fold. The huge expansion in the monetary base, however, was not accompanied by similar increases in *M*1 and *M*2.

Data from: U.S. Federal Reserve.

This huge increase in the monetary base is attributable to actions the Federal Reserve took during the financial crisis and economic downturn of this period. With the financial markets in turmoil, the Fed pursued its job as a lender of last resort with historic vigor. It began by buying large quantities of mortgage-backed securities. Its goal was to restore order to the mortgage market so that would-be homeowners could borrow. Later, the Fed pursued a policy of buying long-term government bonds to keep their prices up and long-term interest rates down. This policy, called *quantitative easing*, is a kind of open-market operation. But rather than buying short-term Treasury bills, as the Fed normally does in an open-market operation, it bought longer-term and somewhat riskier securities. These open-market purchases led to the substantial increase in the monetary base.

The huge expansion in the monetary base, however, did not lead to a similar increase in broader measures of the money supply. While the monetary base increased about 400 percent from 2007 to 2014, *M*1 increased by only 100 percent and *M*2 by only 55 percent. These figures show that the tremendous expansion in the monetary base was accompanied by a large decline in the money multiplier. Why did this decline occur?

The model of the money supply presented earlier in this chapter shows that a key determinant of the money multiplier is the reserve ratio *rr*. From 2007 to 2014, the reserve ratio increased substantially because banks chose to hold substantial quantities of excess reserves. That is, rather than making loans, the banks kept much of their available funds in reserve. (Excess reserves rose from about $1.5 *billion* in 2007 to about $2.5 *trillion* in 2014.) This decision prevented the normal process of money creation that occurs in a system of fractional-reserve banking.

Why did banks choose to hold so much in excess reserves? Part of the reason is that banks had made many bad loans leading up to the financial crisis; when this fact became apparent, bankers tried to tighten their credit standards and make loans only to those they were confident could repay. In addition, interest rates had fallen to such low levels that making loans was not as profitable as it normally is. Banks did not lose much by leaving their financial resources idle as excess reserves.

Although the explosion in the monetary base did not lead to a similar explosion in the money supply, some observers feared that it still might. As the economy recovered from the downturn and interest rates rose to normal levels, they argued, banks could reduce their holdings of excess reserves by making loans. The money supply would start growing, perhaps too quickly.

Policymakers at the Federal Reserve, however, were aware of this potential problem and were ready to handle it. From 2014 to 2017, the Fed increased the interest rate it pays on reserves from 0.25 to 1.50 percent. A higher interest rate on reserves makes holding reserves more profitable for banks, thereby discouraging bank lending and keeping the money multiplier low.[6] ■

Problems in Monetary Control

The Fed has substantial power to influence the money supply, but it cannot control the money supply perfectly. Banks' discretion in how they conduct their businesses, as well as households' decisions about their personal financial affairs, can cause the money supply to change in ways the Fed did not anticipate. For example, if banks choose to hold more excess reserves, the reserve–deposit ratio increases and the money supply falls. Similarly, if households decide to hold more of their money in the form of currency, the currency–deposit ratio increases and the money supply falls. Hence, the money supply sometimes moves in ways the Fed does not intend.

CASE STUDY

Bank Failures and the Money Supply in the 1930s

Between August 1929 and March 1933, the money supply fell 28 percent. As we will discuss in Chapter 12, some economists believe that this large decline in the money supply was the main cause of the Great Depression of the 1930s, when unemployment reached unprecedented levels, prices fell precipitously, and economic hardship was widespread. In light of this hypothesis, one is drawn to ask why the money supply fell so dramatically.

The three variables that determine the money supply—the monetary base, the reserve–deposit ratio, and the currency–deposit ratio—are shown in Table 4-2 for 1929 and 1933. You can see that the fall in the money supply cannot be attributed to a fall in the monetary base: in fact, the monetary base rose 18 percent over this period. Instead, the money supply fell because the money multiplier fell 38 percent. The money multiplier fell because the currency–deposit and reserve–deposit ratios both rose substantially.

TABLE 4-2 The Money Supply and Its Determinants: 1929 and 1933

	August 1929	March 1933
Money Supply	**26.5**	**19.0**
Currency	3.9	5.5
Demand deposits	22.6	13.5
Monetary Base	**7.1**	**8.4**
Currency	3.9	5.5
Reserves	3.2	2.9
Money Multiplier	**3.7**	**2.3**
Reserve–deposit ratio	0.14	0.21
Currency–deposit ratio	0.17	0.41

Data from: Milton Friedman and Anna Schwartz, *A Monetary History of the United States, 1867–1960* (Princeton, NJ: Princeton University Press, 1963), Appendix A.

Most economists attribute the fall in the money multiplier to the large number of bank failures in the early 1930s. From 1930 to 1933, more than 9,000 banks suspended operations, often defaulting on their depositors. The bank failures caused the money supply to fall by altering the behavior of both depositors and bankers.

Bank failures raised the currency–deposit ratio by reducing public confidence in the banking system. People feared that bank failures would continue, and they began to view currency as a more desirable form of money than demand deposits. When they withdrew their deposits, they drained the banks of reserves. The process of money creation reversed itself, as banks responded to lower reserves by reducing their outstanding balance of loans.

In addition, the bank failures raised the reserve–deposit ratio by making bankers more cautious. Having just observed many bank runs, bankers became apprehensive about operating with a small amount of reserves. They therefore increased their holdings of reserves to well above the legal minimum. Just as households responded to the banking crisis by holding more currency relative to deposits, bankers responded by holding more reserves relative to loans. Together these changes caused a large fall in the money multiplier.

Although it is easy to explain why the money supply fell, it is more difficult to decide whether to blame the Federal Reserve. One might argue that the monetary base did not fall, so the Fed should not be blamed. Critics of Fed policy during this period make two counterarguments. First, they claim that the Fed should have taken a more vigorous role in preventing bank failures by acting as a lender of last resort when banks needed cash during bank runs. This would have helped maintain confidence in the banking system and prevented the large fall in the money multiplier. Second, they point out that the Fed could have responded to the fall in the money multiplier by increasing the monetary base even more than it did. Either of these actions would likely have prevented such a large fall in the money supply, which might have reduced the severity of the Great Depression.

Since the 1930s, many policies have been enacted that make such a large and sudden fall in the money supply less likely today. Most important, the system of federal deposit insurance protects depositors when a bank fails. This policy is designed to maintain public confidence in the banking system and thus prevents large swings in the currency–deposit ratio. Deposit insurance has a cost: in the late 1980s and early 1990s, for example, the federal government incurred the large expense of bailing out many insolvent savings-and-loan institutions. Yet deposit insurance helps stabilize the banking system and the money supply. That is why, during the financial crisis of 2008–2009, the Federal Deposit Insurance Corporation raised the amount guaranteed from $100,000 to

$250,000 per depositor. ■

4-4 Conclusion

You should now understand what money is and how central banks affect its supply. Yet this accomplishment, valuable as it is, is only the first step toward understanding monetary policy. The next and more interesting step is to see how changes in the money supply influence the economy. We begin our study of that question in the next chapter. As we examine the effects of monetary policy, we move toward an appreciation of what central bankers can do to improve the functioning of the economy and, just as important, an appreciation of what they cannot do. But be forewarned: you will have to wait until the end of the book to see all the pieces of the puzzle fall into place.

CHAPTER 5

Inflation: Its Causes, Effects, and Social Costs

> Lenin is said to have declared that the best way to destroy the Capitalist System was to debauch the currency. . . . Lenin was certainly right. There is no subtler, no surer means of overturning the existing basis of society than to debauch the currency. The process engages all the hidden forces of economic law on the side of destruction, and does it in a manner which not one man in a million is able to diagnose.
>
> —John Maynard Keynes

In 1970 the *New York Times* cost 15 cents, the median price of a single-family home was $23,400, and the average wage for production workers was $3.39 per hour. In 2017 the *Times* cost $2.50, the median price of a home was $317,200, and the average wage was $20.90 per hour. This overall increase in prices is called **inflation**, which is the subject of this chapter.

The rate of inflation—the percentage change in the overall level of prices—varies greatly over time and across countries. In the United States, according to the consumer price index (CPI), prices rose at an average annual rate of 2.3 percent in the 1960s, 7.1 percent in the 1970s, 5.6 percent in the 1980s, 3.0 percent in the 1990s, and 2.2 percent from 2000 to 2016. Even when the U.S. inflation problem became severe during the 1970s, however, it was nothing compared to the episodes of extraordinarily high inflation, called **hyperinflation**, that other countries have experienced from time to time. A classic example is Germany in 1923, when prices increased an average of 500 percent *per month*. More recently, similar examples of extraordinary inflation gripped the nations of Zimbabwe in 2008 and Venezuela in 2017.

In this chapter we examine the classical theory of the causes, effects, and social costs of inflation. The theory is "classical" in the sense that it assumes that prices are flexible. As discussed in Chapter 1, most economists believe this assumption describes the behavior of the economy in the long run. By contrast, many prices are thought to be sticky in the short run, and beginning in Chapter 10 we incorporate this fact into our analysis. For now, we ignore short-run price stickiness. As we will see, the classical theory of inflation provides a good description of the long run and a useful foundation for the short-run analysis we develop later.

The “hidden forces of economic law” that cause inflation are not as mysterious as Keynes claims in the quotation that opens this chapter. Inflation is simply an increase in average prices, and a price is the rate at which money is exchanged for a good or a service. To understand inflation, we must understand money—what it is, what affects its supply and demand, and what influence it has on the economy. In the previous chapter, we introduced the economist’s concept of “money” and discussed how, in most modern economies, a central bank set up by the government controls the quantity of money in the hands of the public. This chapter begins in Section 5-1 by showing that the quantity of money determines the price level and that the rate of growth in the quantity of money determines the rate of inflation.

Inflation in turn has numerous effects of its own on the economy. Section 5-2 discusses the revenue that governments can raise by printing money, sometimes called the *inflation tax.* Section 5-3 examines how inflation affects the nominal interest rate. Section 5-4 discusses how the nominal interest rate affects the quantity of money people wish to hold and, thereby, the price level.

After analyzing the causes and effects of inflation, in Section 5-5 we address what is perhaps the most important question about inflation: Is it a major social problem? Does inflation amount to “overturning the existing basis of society,” as the chapter’s opening quotation suggests?

Finally, in Section 5-6, we discuss the dramatic case of hyperinflation. Hyperinflations are interesting to examine because they show clearly the causes, effects, and costs of inflation. Just as seismologists learn much about plate tectonics by studying earthquakes, economists learn much about money and prices by studying how hyperinflations begin and end.

5-1 The Quantity Theory of Money

In Chapter 4 we defined what money is and learned that the quantity of money available in the economy is called the money supply. We also saw how the money supply is determined by the banking system together with the policy decisions of the central bank. With that foundation, we can now start to examine the macroeconomic effects of monetary policy. To do this, we need a theory that tells us how the quantity of money is related to other economic variables, such as prices and incomes. The theory we develop in this section, called the *quantity theory of money*, has its roots in the work of the early monetary theorists, including the philosopher and economist David Hume (1711–1776). It remains the leading explanation for how money affects the economy in the long run.

Transactions and the Quantity Equation

If you hear an economist use the word "supply," you can be sure that the word "demand" is not far behind. Indeed, having fully explored the supply of money, we now focus on the demand for it.

The starting point of the quantity theory of money is the insight that people hold money to buy goods and services. The more money they need for such transactions, the more money they hold. Thus, the quantity of money in the economy is related to the number of dollars exchanged in transactions.

The link between transactions and money is expressed in the following equation, called the **quantity equation**:

Money×Velocity=Price×TransactionsM×V=P×T.

$$\begin{array}{ccccccc} \text{Money} & \times & \text{Velocity} & = & \text{Price} & \times & \text{Transactions} \\ M & \times & V & = & P & \times & T. \end{array}$$

Let's examine each of the four variables in this equation.

The right-hand side of the quantity equation tells us about transactions. *T* represents the total number of transactions during some period of time, say, a year. In other words, *T* is the number of times in a year that goods or services are exchanged for money. *P* is the price of a typical transaction—the number of dollars exchanged. The product of the price of a transaction and the number of transactions, *PT*, equals the number of dollars exchanged in a year.

The left-hand side of the quantity equation tells us about the money used to make the transactions. *M* is the quantity of money. *V*, called the **transactions velocity of money**, measures the rate at which money circulates in the economy. In other words, velocity tells us the number of times a dollar bill changes hands in a given period of time.

For example, suppose that 50 loaves of bread are sold in a given year at $2 per loaf. Then *T* equals 50 loaves per year, and *P* equals $2 per loaf. The total number of dollars exchanged is

PT=$2/loaf×50 loaves/year=$100/year.

$$PT = \$2/\text{loaf} \times 50 \text{ loaves}/\text{year} = \$100/\text{year}.$$

The right-hand side of the quantity equation equals $100 per year, the dollar value of all transactions.

Suppose further that the quantity of money in the economy is $20. By rearranging the quantity equation, we can compute velocity as

V=PT/M=($100/year)/($20)=5 times per year.

$$\begin{aligned} V &= PT/M \\ &= (\$100/\text{year}) \,/\, (\$20) \\ &= 5 \text{ times per year.} \end{aligned}$$

That is, for $100 of transactions per year to take place with $20 of money, each dollar must change hands 5 times per year.

The quantity equation is an *identity*: the definitions of the four variables make it true. This type of equation is useful because it shows that if one of the variables changes, one or more of the others must also change to maintain the equality. For example, if the quantity of money increases and the velocity of money remains constant, then either the price or the number of transactions must rise.

From Transactions to Income

When studying the role of money in the economy, economists usually use a slightly different version of the quantity equation than the one just introduced. The problem with the first equation is that the number of transactions is difficult to measure. To solve this problem, the number of transactions *T* is replaced by the total output of the economy *Y*.

Transactions and output are related because the more the economy produces, the more goods are bought and sold. Yet they are not the same. When one person sells a used car to another person, for example, they make a transaction using money, even though the used car is not part of current output. Nonetheless, the dollar value of transactions is roughly proportional to the dollar value of output.

If *Y* denotes the amount of output and *P* denotes the price of one unit of output, then the dollar value of output is *PY*. We encountered measures for these variables when we discussed the national income accounts in Chapter 2: *Y* is real GDP; *P*, the GDP deflator; and *PY*, nominal GDP. The quantity equation becomes

$$\begin{array}{ccccccc} \text{Money} & \times & \text{Velocity} & = & \text{Price} & \times & \text{Output} \\ M & \times & V & = & P & \times & Y. \end{array}$$

Because *Y* is also total income, *V* in this version of the quantity equation is called the **income velocity of money**. The income velocity of money tells us the number of times a dollar bill enters someone's income in a given period of time. This version of the quantity equation is the most common, and it is the one we use from now on.

The Money Demand Function and the Quantity Equation

When we analyze how money affects the economy, it is often useful to express the quantity of money in terms of the quantity of goods and services it can buy. This amount, *M/P*, is called **real money balances**.

Real money balances measure the purchasing power of the stock of money. For example, consider an economy that produces only bread. If the quantity of money is $20, and the price of a loaf is $2, then real money balances are 10 loaves of bread. That is, at current prices, the stock of money in the economy can buy 10 loaves.

A **money demand function** is an equation that shows the determinants of the quantity of real money balances people wish to hold. A simple money demand function is

$$(M/P)^d = kY,$$

where *k* is a constant that tells us how much money people want to hold for every dollar of income. This equation states that the quantity of real money balances demanded is proportional to real income.

The money demand function is like the demand function for a particular good. Here the "good" is the convenience of holding real money balances. Just as owning an automobile makes it easier for a person to travel, holding money makes it easier to make transactions. Therefore, just as higher income leads to a greater demand for automobiles, higher income also leads to a greater demand for real money balances.

This money demand function offers another way to view the quantity equation. To see this, add to the money demand function the condition that the demand for real money balances $(M/P)^d$ must equal the supply M/P. Therefore,

$$M/P = kY.$$

A simple rearrangement of terms changes this equation into

$$M\,(1/k) = PY,$$

which can be written as

$$MV = PY,$$

where $V = 1/k$. These few steps of simple mathematics show the link between the demand for money and the velocity of money. When people want to hold a lot of money for each dollar of income (*k* is large), money changes hands infrequently (*V* is small). Conversely, when people want to hold only a little money (*k* is small), money changes hands frequently (*V* is large). In other words, the money demand parameter *k* and the velocity of money *V* are opposite sides of the same coin.

The Assumption of Constant Velocity

The quantity equation can be viewed as a definition: it defines velocity *V* as the ratio of nominal GDP, *PY*, to the quantity of money *M*. Yet if we make the additional assumption that the velocity of money is constant, then the quantity equation becomes a useful theory about the effects of money, called the **quantity theory of money**.

Like many of the assumptions in economics, the assumption of constant velocity is only a simplification of reality. Velocity does change if the money demand function changes. For example, when automatic teller machines were introduced, people could reduce their average money holdings, which meant a fall in the money demand parameter *k* and an increase in velocity *V*. Nonetheless, experience shows that the assumption of constant velocity is useful in many situations. Let's therefore assume that velocity is constant and see what this assumption implies about the effects of the money supply on the economy.

With this assumption included, the quantity equation can be seen as a theory of what determines nominal GDP. The quantity equation says

$$M\overline{V} = PY,$$

where the bar over *V* means that velocity is fixed. Therefore, a change in the quantity of money (*M*) must cause a proportionate change in nominal GDP (*PY*). That is, if velocity is fixed, the quantity of money determines the dollar value of the economy's output.

Money, Prices, and Inflation

We now have a theory to explain what determines the economy's level of prices. The theory has three building blocks:

1. The factors of production and the production function determine output *Y*. We borrow this conclusion from Chapter 3.
2. The money supply *M* set by the central bank determines the nominal value of output *PY*. This conclusion follows from the quantity equation and the assumption that the velocity of money is fixed.
3. The price level *P* is then the ratio of the nominal value of output *PY* to output *Y*.

In other words, the productive capability of the economy determines real GDP, the quantity of money determines nominal GDP, and the GDP deflator is the ratio of nominal GDP to real GDP.

This theory explains what happens when the central bank changes the supply of money. Because velocity *V* is fixed, any change in the money supply *M* must lead to a proportionate change in the nominal value of output *PY*. Because the factors of production and the production function have already determined output *Y*, the nominal value of output *PY* can adjust only if the price level *P* changes. Hence, the quantity theory implies that the price level is proportional to the money supply.

Because the inflation rate is the percentage change in the price level, this theory of the price level is also a theory of the inflation rate. The quantity equation, written in percentage-change form, is

$$\%\Delta M + \%\Delta V = \%\Delta P + \%\Delta Y.$$

Consider each of these four terms. First, the percentage change in the quantity of money, $\%\Delta M$, is under the control of the central bank. Second, the percentage change in velocity, $\%\Delta V$, reflects shifts in money demand; we have assumed that velocity is constant, so $\%\Delta V$ is zero. Third, the percentage change in the price level, $\%\Delta P$, is the rate of inflation; this is the variable in the equation that we would like to explain. Fourth, the percentage change in output, $\%\Delta Y$, depends on growth in the factors of production and on technological progress, which for our present purposes we are taking as given. This analysis tells us that (except for a constant that depends on exogenous growth in output) the growth in the money supply determines the rate of inflation.

Thus, the quantity theory of money states that the central bank, which controls the money supply, has ultimate control over the rate of inflation. If the central bank keeps the money supply stable, the price level will be stable. If the central bank increases the money supply rapidly, the price level will rise rapidly.

CASE STUDY

Inflation and Money Growth

“Inflation is always and everywhere a monetary phenomenon.” So wrote Milton Friedman, the great economist who won the Nobel Prize in economics in 1976. The quantity theory of money leads us to agree that the growth in the quantity of money is the primary determinant of the inflation rate. Yet Friedman’s claim is empirical, not theoretical. To evaluate his claim, and to judge the usefulness of our theory, we need to look at data on money and prices.

Friedman, together with fellow economist Anna Schwartz, wrote two treatises on monetary history that documented the sources and effects of changes in the quantity of money over the past century.[1] Figure 5-1 uses their data and plots the average rate of money growth and the average rate of inflation in the United States over each decade since the 1870s. The data confirm the link between inflation and growth in the quantity of money. Decades with high money growth (such as the 1970s) tend to have high inflation, and decades with low money growth (such as the 1930s) tend to have low inflation.

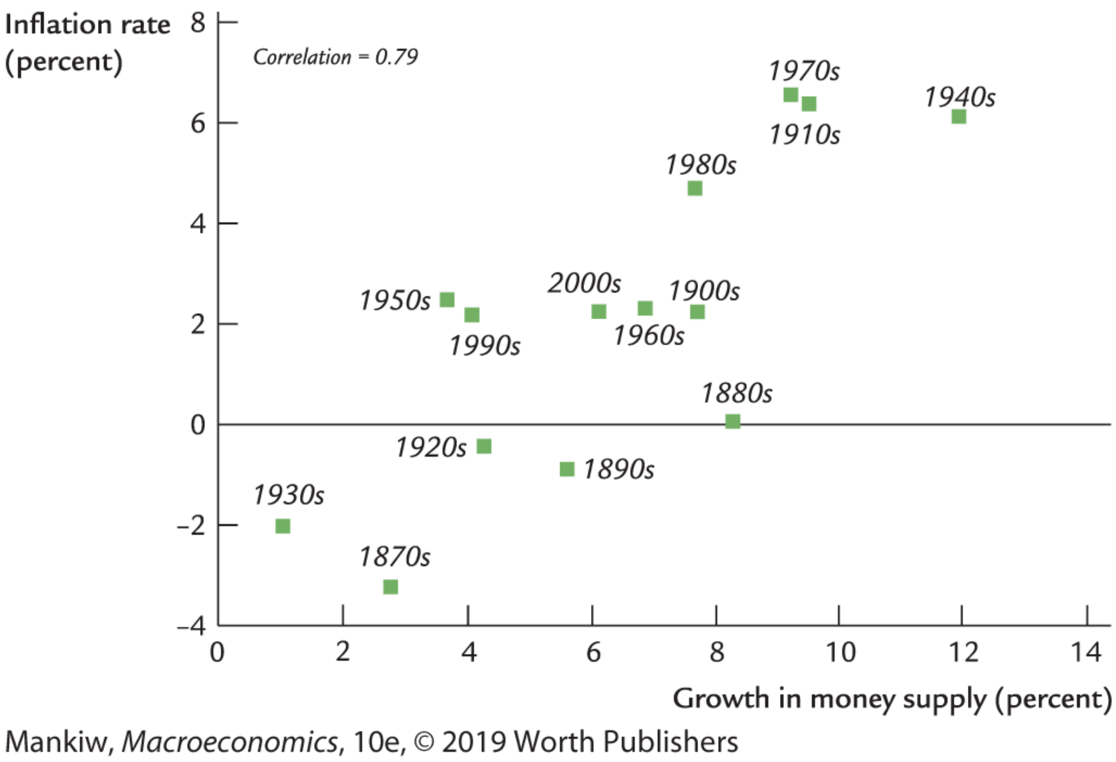
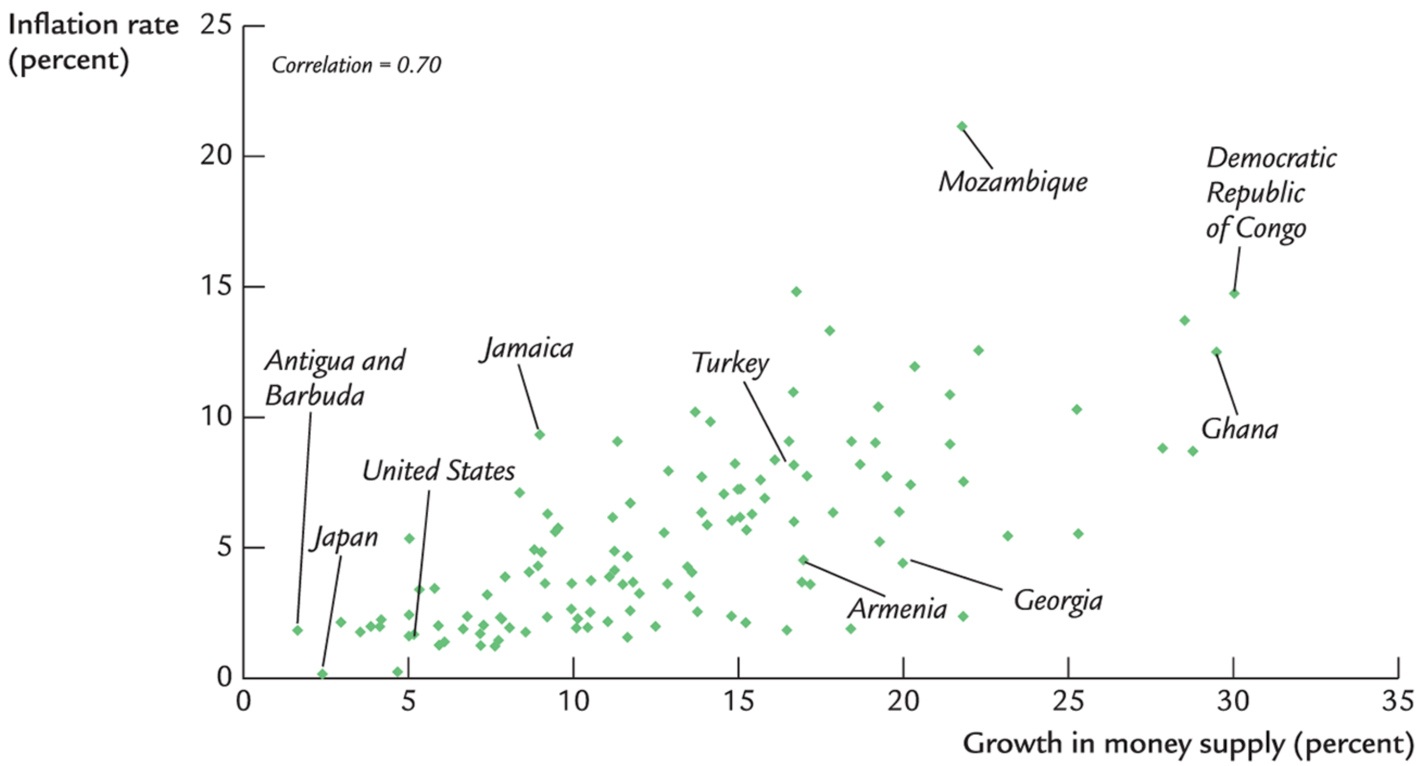


FIGURE 5-1 Historical Data on U.S. Inflation and Money Growth In this scatterplot of money growth and inflation, each point represents a decade. The horizontal axis shows the average growth in the money supply (as measured by *M*2) over the decade, and the vertical axis shows the average rate of inflation (as measured by the GDP deflator). The positive correlation between money growth and inflation is evidence for the quantity theory's prediction that high money growth leads to high inflation.

Data from: For the data through the 1960s: Milton Friedman and Anna J. Schwartz, *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates 1867–1975* (Chicago: University of Chicago Press, 1982). For recent data: U.S. Department of Commerce and Federal Reserve Board.

As you may have learned in a statistics class, one way to quantify a relationship between two variables is with a measure called *correlation*. A correlation is +1 if the two variables move exactly in tandem, 0 if they are unrelated, and –1 if they move exactly opposite each other. In Figure 5-1, the correlation is 0.79, indicating that the two variables move closely together.

Figure 5-2 examines the same question using international data. It shows the average rate of inflation and the average rate of money growth in 123 countries during the period from 2007 to 2016. Again, the link between money growth and inflation is clear. Countries with high money growth (such as Ghana and Mozambique) tend to have high inflation, and countries with low money growth (such as Japan and the United States) tend to have low inflation. The correlation here is 0.70.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 5-2 International Data on Inflation and Money Growth In this scatterplot, each point represents a country. The horizontal axis shows the average growth in the money supply (as measured by a broad monetary aggregate) from 2007 to 2016, and the vertical axis shows the average rate of inflation (as measured by the CPI). Once again, the positive correlation is evidence for the quantity theory's prediction that high money growth leads to high inflation.

Data from: International Monetary Fund.

If we looked at monthly data on money growth and inflation, rather than data for decade-long periods, we would not see as close a connection between these two variables. This theory of inflation works best in the long run, not in the short run. We examine the short-run impact of changes in the quantity of money when we turn to economic fluctuations in Part Four of this book. ■

5-2 Seigniorage: The Revenue from Printing Money

So far, we have seen how growth in the money supply causes inflation. With inflation as a consequence, what would ever induce a central bank to increase the money supply substantially? Here we examine one answer to this question.

Let's start with an indisputable fact: all governments spend money. Some of this spending is to buy goods and services (such as roads and police), and some is to provide transfer payments (for the poor and elderly, for example). A government can finance its spending in three ways. First, it can raise revenue through taxes, such as personal and corporate income taxes. Second, it can borrow from the public by selling government bonds. Third, it can print money.

The revenue raised by the printing of money is called **seigniorage**. The term comes from *seigneur*, the French word for "feudal lord." In the Middle Ages, the lord had the exclusive right on his manor to coin money. Today this right belongs to the central government, and it is one source of revenue.

When the government prints money to finance expenditure, it increases the money supply. The increase in the money supply, in turn, causes inflation. Printing money to raise revenue is like imposing an *inflation tax.*

At first, inflation might not look like a tax. After all, no one receives a bill for it—the government just prints the money it needs. Who, then, pays the inflation tax? The answer is the holders of money. As prices rise, the real value of the money in your wallet falls. Therefore, when the government prints new money for its use, it makes the old money in the hands of the public less valuable. In essence, inflation is a tax on holding money.

The amount of revenue raised by printing money varies from country to country. In the United States, the amount has been small: seigniorage has usually accounted for less than 3 percent of government revenue. In Italy and Greece, seigniorage has often been more than 10 percent of government revenue.[2] In countries experiencing hyperinflation, seigniorage is often the government's chief source of revenue—indeed, the need to print money to finance expenditure is a primary cause of hyperinflation.

CASE STUDY

Paying for the American Revolution

Although seigniorage has not been a major source of revenue for the U.S. government in recent history, the situation was very different two and a half centuries ago. Beginning in 1775, the Continental Congress needed to find a way to finance the Revolution, but it had limited ability to raise revenue through taxation. It therefore relied on the printing of fiat money to help pay for the war.

The Continental Congress's reliance on seigniorage increased over time. New issues of continental currency were about $6 million in 1775, $19 million in 1776, and $13 million in 1777. This amount increased to $63 million in 1778 and $125 million in 1779.

Not surprisingly, this rapid growth in the money supply led to massive inflation. At the end of the war, the price of gold measured in continental dollars was more than 100 times its level of only a few years earlier. The large quantity of the continental currency made the continental dollar nearly worthless. This experience also gave birth to a once-popular expression: people used to say something was "not worth a continental" to mean that the item had little real value.

When the new nation won its independence, there was a natural skepticism about fiat money. Upon the recommendation of the first secretary of the Treasury, Alexander Hamilton, Congress passed the Mint Act of 1792, which established gold and silver as the basis for a new system of commodity money. ■

5-3 Inflation and Interest Rates

As we first discussed in Chapter 3, interest rates are among the most important macroeconomic variables. They are the prices that link the present and the future. Here we discuss the relationship between inflation and interest rates.

Two Interest Rates: Real and Nominal

Suppose you deposit your savings in a bank account that pays 8 percent interest annually. Next year, you withdraw your savings and the accumulated interest. Are you 8 percent richer than you were when you made the deposit a year earlier?

The answer depends on what “richer” means. To be sure, you have 8 percent more dollars than you had before. But if prices have risen, each dollar buys less, and your purchasing power has not risen by 8 percent. If the inflation rate was 5 percent over the year, then the amount of goods you can buy has increased by only 3 percent. And if the inflation rate was 10 percent, then your purchasing power has fallen by 2 percent.

The interest rate that the bank pays is the **nominal interest rate**, and the increase in your purchasing power is the **real interest rate**. If i denotes the nominal interest rate, r the real interest rate, and π the rate of inflation, the relationship among these three variables can be written as

$$r = i - \pi.$$

The real interest rate is the difference between the nominal interest rate and the rate of inflation.[3]

The Fisher Effect

Rearranging terms in our equation for the real interest rate, we can show that the nominal interest rate is the sum of the real interest rate and the inflation rate:

$$i = r + \pi.$$

The equation written in this way is called the **Fisher equation**, after economist Irving Fisher (1867–1947). It shows that the nominal interest rate can change for two reasons: because the real interest rate changes or because the inflation rate changes.

Once we separate the nominal interest rate into these two parts, we can use this equation to develop a theory that explains the nominal interest rate. Chapter 3 showed that the real interest rate adjusts to equilibrate saving and investment. The quantity theory of money shows that the rate of money growth determines the rate of inflation. The Fisher equation then tells us to add the real interest rate and the inflation rate together to determine the nominal interest rate.

The quantity theory and the Fisher equation together tell us how money growth affects the nominal interest rate. *According to the quantity theory, an increase in the rate of money growth of 1 percent causes a 1 percent increase in the rate of inflation. According to the Fisher equation, a 1 percent increase in the rate of inflation in turn causes a 1 percent increase in the nominal interest rate.* The one-for-one relation between the inflation rate and the nominal interest rate is called the **Fisher effect**.

CASE STUDY

Inflation and Nominal Interest Rates

How useful is the Fisher effect in explaining interest rates? To answer this question, we look at two types of data on inflation and nominal interest rates.

Figure 5-3 shows the variation over time in the nominal interest rate and the inflation rate in the United States from 1954 to 2016. You can see that the Fisher effect has done a good job of explaining fluctuations in the nominal interest rate during this period. When inflation is high, nominal interest rates are typically high, and when inflation is low, nominal interest rates are typically low as well. The correlation between the inflation rate and the nominal interest rate is 0.76.

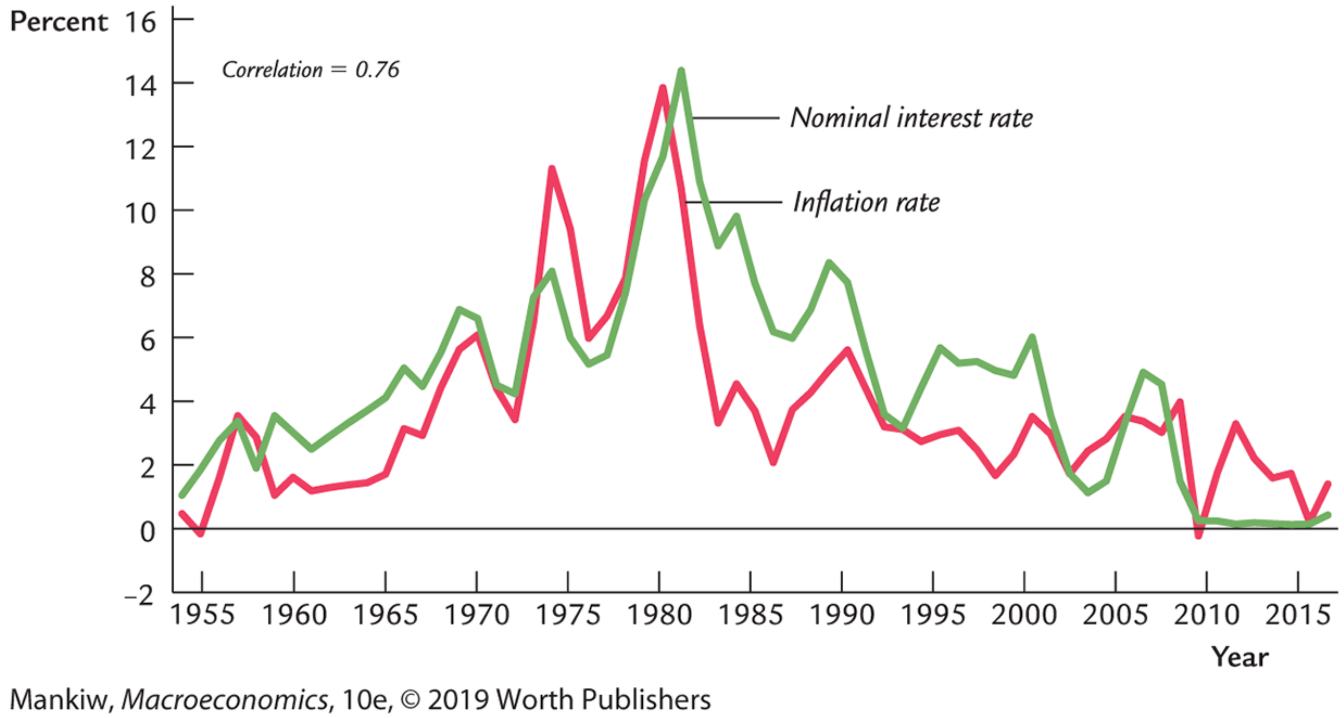


FIGURE 5-3 Inflation and Nominal Interest Rates Over Time This figure plots the nominal interest rate (on three-month Treasury bills) and

the inflation rate (as measured by the CPI) in the United States since 1954. It shows the Fisher effect: higher inflation leads to a higher nominal interest rate.

Data from: Federal Reserve.

Similar support for the Fisher effect comes from examining the variation across countries. As Figure 5-4 shows, a nation’s inflation rate and its nominal interest rate are related. Countries with high inflation tend to have high nominal interest rates as well, and countries with low inflation tend to have low nominal interest rates. The correlation between these two variables is 0.75.

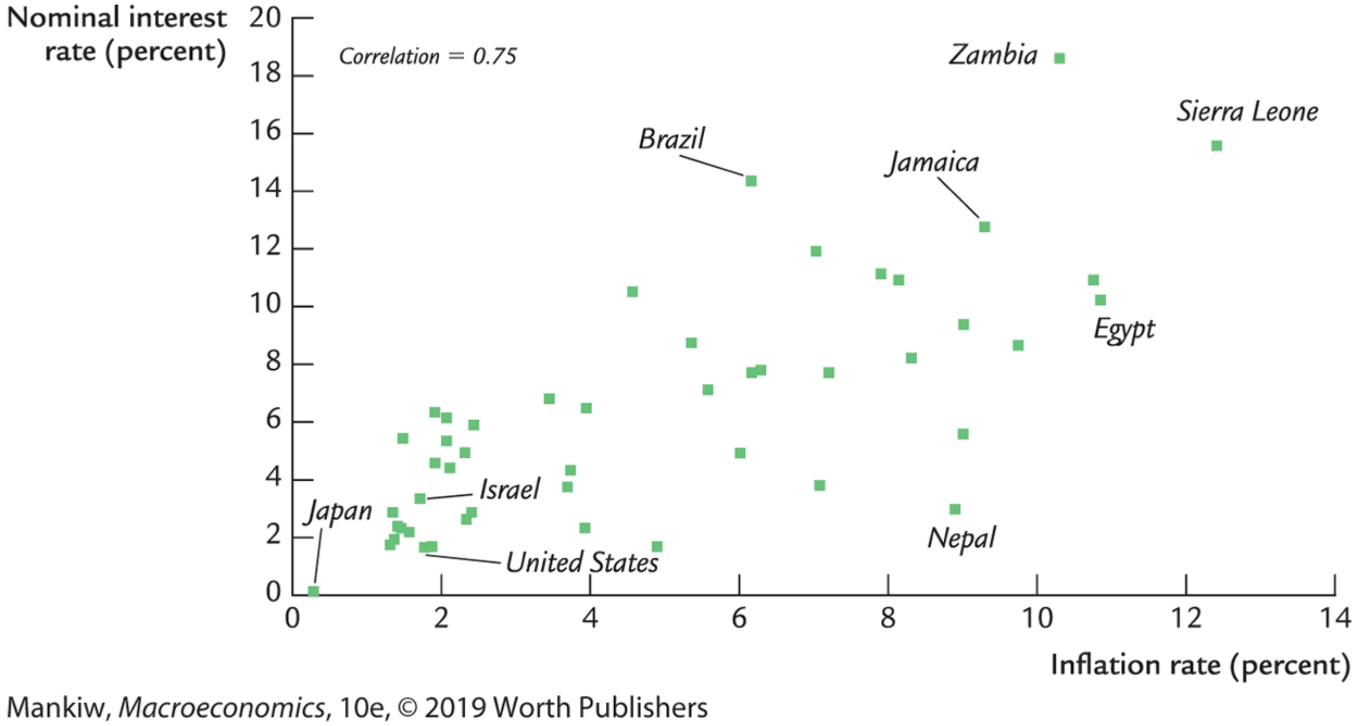


FIGURE 5-4 Inflation and Nominal Interest Rates Across Countries This scatterplot shows the average nominal interest rate on short-term Treasury bills and the average inflation rate in 48 countries from 2007 to 2016. The positive correlation between the inflation rate and the nominal interest rate is evidence for the Fisher effect.

Data from: International Monetary Fund.

The link between inflation and nominal interest rates is well known to Wall Street investment firms. Because bond prices move inversely with interest rates, one can get rich by correctly predicting the direction in which interest rates will move. Many Wall Street firms hire *Fed watchers* to monitor monetary policy and news about inflation to anticipate changes in interest rates. ■

Two Real Interest Rates: *Ex Ante* and *Ex Post*

When a borrower and lender agree on a nominal interest rate, they do not know what the inflation rate over the term of the loan will be. Therefore, we must distinguish between two concepts of the real interest rate: the real interest rate that the borrower and lender expect when the loan is made, called the ***ex ante* real interest rate**, and the real interest rate that is actually realized, called the ***ex post* real interest rate**.

Although borrowers and lenders cannot predict future inflation with certainty, they do have some expectation about what the inflation rate will be. Let π denote actual future inflation and $E\pi$ denote expected future inflation. The *ex ante* real interest rate is $i - E\pi$, and the *ex post* real interest rate is $i - \pi$. The two real interest rates differ when actual inflation π differs from expected inflation $E\pi$.

How does this distinction between actual and expected inflation modify the Fisher effect? Clearly, the nominal interest rate cannot adjust to actual inflation, because actual inflation is not known when the nominal interest rate is set. The nominal interest rate can adjust only to expected inflation. The Fisher effect is more precisely written as

$$i = r + E\pi.$$

The *ex ante* real interest rate *r* is determined by equilibrium in the market for goods and services, as described by the model in Chapter 3. The nominal interest rate *i* moves one-for-one with changes in expected inflation $E\pi$.

If the nominal interest rate is supposed to respond to expected inflation, why do we see such as a strong correlation between nominal interest rates and actual inflation in Figures 5-3 and 5-4? The reason is that actual inflation is usually persistent and, therefore, high actual inflation goes along with high expected inflation. But that need not always be the case. During the late nineteenth and early twentieth centuries, inflation showed little persistence. When people experienced high inflation, they had no reason to expect high inflation to continue. As a result, the correlation between nominal interest rates and actual inflation was much weaker. Fisher himself noted this fact and suggested that inflation “caught merchants napping.”[4]

5-4 The Nominal Interest Rate and the Demand for Money

The quantity theory is based on a simple money demand function: it assumes that the demand for real money balances is proportional to income. The quantity theory is a good place to start when analyzing the effects of money, but it is not the whole story. Here we add another determinant of the quantity of money demanded—the nominal interest rate.

The Cost of Holding Money

The money you hold in your wallet does not earn interest. If, instead of holding that money, you used it to buy government bonds or deposited it in a savings account, you would earn the nominal interest rate. Therefore, the nominal interest rate is the opportunity cost of holding money: it is what you give up by holding money rather than bonds.

Another way to see that the cost of holding money equals the nominal interest rate is by comparing the real returns on alternative assets. Assets other than money, such as government bonds, earn the real return *r*. Money earns an expected real return of $-E\pi$, because its real value declines at the rate of inflation. When you hold money, you give up the difference between these two returns. Thus, the cost of holding money is $r - (-E\pi)$, which the Fisher equation tells us is the nominal interest rate *i*.

Just as the quantity of bread demanded depends on the price of bread, the quantity of money demanded depends on the price of holding money. Hence, the demand for real money balances depends on both income and the nominal interest rate. We write the general money demand function as

$$(M/P)^d = L(i, Y).$$

The letter *L* is used to denote money demand because money is the economy's most liquid asset (the asset most easily used to make transactions). This equation states that the demand for the liquidity of real money balances is a function of income and the nominal interest rate. The higher the level of income *Y*, the greater the demand for real money balances. The higher the nominal interest rate *i*, the lower the demand for real money balances.

Future Money and Current Prices

Money, prices, and interest rates are now related in several ways. Figure 5-5 illustrates the linkages we have discussed. As the quantity theory of money explains, money supply and money demand together determine the equilibrium price level. Changes in the price level are, by definition, the rate of inflation. Inflation, in turn, affects the nominal interest rate through the Fisher effect. But now, because the nominal interest rate is the cost of holding money, the nominal interest rate feeds back to affect the demand for money.

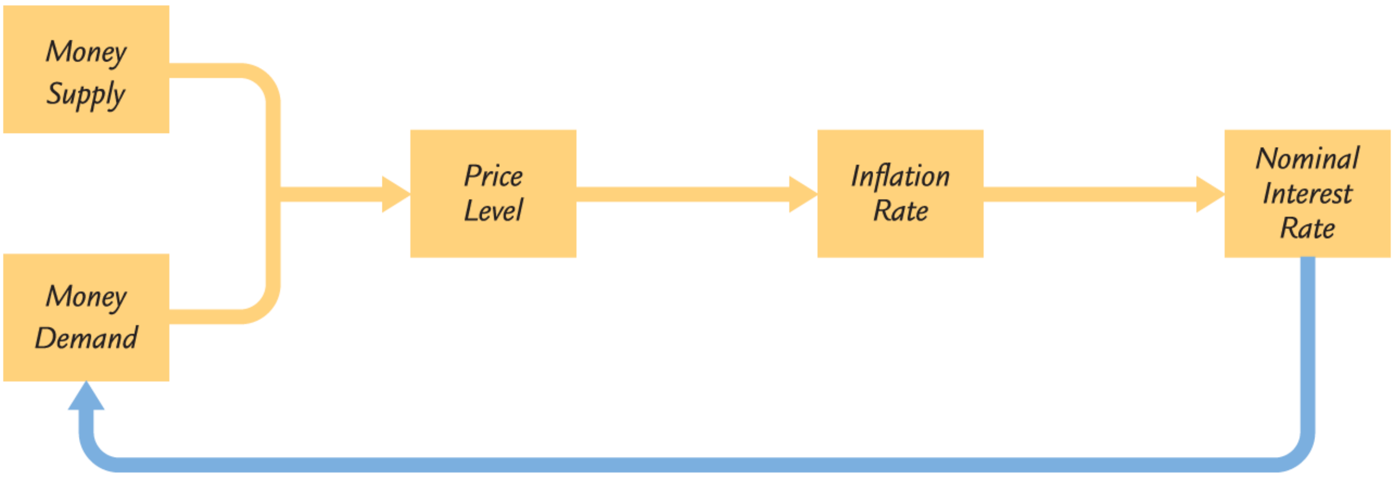


FIGURE 5-5 The Linkages Among Money, Prices, and Interest Rates This figure illustrates the relationships among money, prices, and interest rates. Money supply and money demand determine the price level. Changes in the price level determine the inflation rate. The inflation rate influences the nominal interest rate. Because the nominal interest rate is the cost of holding money, it may affect money demand. This last link (shown as a blue line) is omitted from the basic quantity theory of money.

Consider how the introduction of this last link affects our theory of the price level. First, equate the supply of real money balances M/P to the demand *L(i, Y):*

$$M/P = L(i, Y).$$

Next, use the Fisher equation to write the nominal interest rate as the sum of the real interest rate and expected inflation:

$$M/P = L(r + E\pi, Y).$$

This equation states that the level of real money balances depends on the expected rate of inflation.

The last equation tells a more sophisticated story about the determination of the price level than does the quantity theory. The quantity theory of money says that today's money supply determines today's price level. This conclusion remains partly true: if the nominal interest rate and output are held constant, the price level moves proportionately with the money supply. Yet the nominal interest rate is not constant; it depends on expected inflation, which in turn depends on growth in the money supply. The presence of the nominal interest rate in the money demand function yields an additional channel through which money supply affects the price level.

This general money demand equation implies that the price level depends not only on today's money supply but also on the money supply expected in the future. To see why, suppose the Fed announces that it will increase the money supply in the future, but it does not change the money supply today. This announcement causes people to expect higher money growth and higher inflation. Through the Fisher effect, this increase in expected inflation raises the nominal interest rate. The higher nominal interest rate increases the cost of holding money and therefore reduces the demand for real money balances. Because the Fed has not changed the quantity of money available today, the reduced demand for real money balances leads to a higher price level. Hence, expectations of higher money growth in the future lead to a higher price level today.

The effect of money on prices is thus more complicated than the simplest quantity theory suggests. Formal models show what determines the price level with a more general money demand function. These models are beyond the scope of this text, but their bottom line is simple. *The price level depends on a weighted average of the current money supply and the money supply expected to prevail in the future. Inflation is driven by both current growth in the money supply and its expected future growth.*

5-5 The Social Costs of Inflation

Our discussion of the causes and effects of inflation does not tell us much about the social problems that result from inflation. We turn to those problems now.

The Layman’s View and the Classical Response

If you ask the average person why inflation is a social problem, she will probably answer that inflation makes her poorer. “Each year my boss gives me a raise, but prices go up and that takes some of my raise away from me.” The implicit assumption in this statement is that if there were no inflation, she would get the same raise and be able to buy more goods.

This complaint about inflation is a common fallacy. As we know from Chapter 3, the purchasing power of labor—the real wage—depends on the marginal productivity of labor, not on how much money the government prints. If the central bank reduces inflation by slowing the rate of money growth, workers will not see their real wages increasing more rapidly. Instead, when inflation slows, firms will increase the prices of their products less each year and, as a result, will give their workers smaller raises.

According to the classical theory of money, a change in the price level is like a change in the units of measurement. It is as if we switched from measuring distances in feet to measuring them in inches: numbers get larger, but nothing really changes. Imagine that tomorrow morning you wake up and find that, for some reason, all dollar figures in the economy have been multiplied by ten. The price of everything you buy has increased 10-fold, but so have your wage and the value of your savings. What difference would such a price increase make to your life? All numbers would have an extra zero at the end, but nothing else would change. Your economic well-being depends on relative prices, not the overall price level.

Why, then, is a persistent increase in the price level a social problem? It turns out that the costs of inflation are subtle. Indeed, economists disagree about the size of the social costs. To the surprise of many laymen, some economists argue that the costs of inflation are small—at least for the moderate rates of inflation that most countries have experienced in recent years.[5]

CASE STUDY

What Economists and the Public Say About Inflation

As we have been discussing, laymen and economists hold very different views about the costs of inflation. In

1996, economist (and 2013 Nobel Prize winner) Robert Shiller documented this difference of opinion in a survey of the two groups. The survey results are striking, for they show how the study of economics changes a person's attitudes.

In one question, Shiller asked people whether their "biggest gripe about inflation" was that "inflation hurts my real buying power, it makes me poorer." Of the general public, 77 percent agreed with this statement, compared to only 12 percent of economists. Shiller also asked people whether they agreed with the following statement: "When I see projections about how many times more a college education will cost, or how many times more the cost of living will be in coming decades, I feel a sense of uneasiness; these inflation projections really make me worry that my own income will not rise as much as such costs will." Among the general public, 66 percent said they fully agreed with this statement, whereas only 5 percent of economists agreed with it.

Survey respondents were asked to judge the seriousness of inflation as a policy problem: "Do you agree that preventing high inflation is an important national priority, as important as preventing drug abuse or preventing deterioration in the quality of our schools?" Shiller found that 52 percent of laymen, but only 18 percent of economists, fully agreed with this view. Apparently, inflation worries the public much more than it does the economics profession.

The public's distaste for inflation may be psychological. Shiller asked those surveyed if they agreed with the following statement: "I think that if my pay went up I would feel more satisfaction in my job, more sense of fulfillment, even if prices went up just as much." Of the public, 49 percent fully or partly agreed with this statement, compared to 8 percent of economists.

Do these survey results mean that laymen are wrong and economists are right about the costs of inflation? Not necessarily. But economists have the advantage of having given the issue more thought. So let's now consider what some of the costs of inflation might be.[6] ■

The Costs of Expected Inflation

Consider first the case of expected inflation. Suppose that every month the price level rose by 1/2 percent. What would be the social costs of such a steady and predictable 6 percent annual inflation?

One cost is the distorting effect of the inflation tax on the amount of money people hold. As we have already discussed, a higher inflation rate leads to a higher nominal interest rate, which in turn leads to lower real money balances. But for people to hold lower money balances and spend the same amount, they must make more frequent trips to the bank to withdraw money—for example, they might withdraw $50 twice a week rather than $100 once a week. The inconvenience of reducing money holding is metaphorically called the **shoeleather cost** of inflation, because walking to the bank more often causes one's shoes to wear out more quickly.

A second cost of inflation arises because high inflation induces firms to change their posted prices more often. Changing prices is sometimes costly; for example, it may require printing and distributing a new catalog. These costs are called **menu costs**, because the higher the rate of inflation, the more often restaurants have to print new menus.

A third cost of inflation arises because firms facing menu costs change prices infrequently; therefore, the higher the rate of inflation, the greater the variability in relative prices. For example, suppose a firm issues a new catalog every January. If there is no inflation, then the firm's prices relative to the overall price level are constant over the year. Yet if inflation is 1/2 percent per month, then from the beginning to the end of the year the firm's relative prices fall by 6 percent. Sales from this catalog will tend to be low early in the year (when its prices are relatively high) and high later in the year (when its prices are relatively low). Hence, when inflation induces variability in relative prices, it leads to microeconomic inefficiencies in the allocation of resources.

A fourth cost of inflation results from the tax laws. Many provisions of the tax code do not take into account the effects of inflation. Inflation can alter individuals' tax liability, often in ways that lawmakers did not intend.

One example of the failure of the tax code to deal with inflation is the tax treatment of capital gains. Suppose you buy some stock today and sell it a year from now at the same real price. It would seem reasonable for the government not to levy a tax, because you have earned no real income from this investment. Indeed, if there is no inflation, a zero tax liability would be the outcome. But suppose the rate of inflation is 6 percent and you initially paid $100 per share for the stock; for the real price to be the same a year later, you must sell the stock for $106 per share. In this case the tax code, which ignores the effects of inflation, says that you have earned $6 per share in income, and the government taxes you on this capital gain. The problem is that the tax code measures income as the nominal rather than the real capital gain. In this example, and in many others, inflation distorts how taxes are levied.

A fifth cost of inflation is the inconvenience of living in a world with a changing price level. Money is the yardstick with which we measure economic transactions. When there is inflation, that yardstick is changing in length. To continue the analogy, suppose that Congress passed a law specifying that a yard would equal 36 inches in 2019, 35 inches in 2020, 34 inches in 2021, and so on. The law would result in no ambiguity but would be highly inconvenient. When someone measured a distance in yards, it would be necessary to specify whether the measurement was in 2020 yards or 2021 yards; to compare distances measured in different years, one would need to make an "inflation" correction. Similarly, the dollar is a less useful measure when its value is always changing. The changing value of the dollar requires that we correct for inflation when comparing dollar figures from different times.

For example, a changing price level complicates personal financial planning. An important decision that all households face is how much of their income to consume today and how much to save for retirement. A dollar saved today and invested at a fixed nominal interest rate will yield a fixed dollar amount in the future. Yet the real value of that dollar amount—which will determine the retiree's living standard—depends on the future price level. Deciding how much to save would be simpler if people could count on the price level in 30 years being similar to its level today.

The Costs of Unexpected Inflation

Unexpected inflation has an effect that is more pernicious than any of the costs of steady, anticipated inflation: it arbitrarily redistributes wealth among people. You can see how this works by examining long-term loans. Most loan agreements specify a nominal interest rate, which is based on the rate of inflation expected at the time of the agreement. If inflation turns out differently from what was expected, the *ex post* real return that the debtor pays to the creditor differs from what both parties anticipated. On the one hand, if inflation turns out to be higher than expected, the debtor wins and the creditor loses because the debtor repays the loan with less valuable dollars. On the other hand, if inflation turns out to be lower than expected, the creditor wins and the debtor loses because the repayment is worth more than the two parties anticipated.

Consider, for example, a person taking out a mortgage in 1960. At the time, a 30-year mortgage had an interest rate of about 6 percent per year. This rate was based on a low rate of expected inflation—inflation over the previous decade had averaged only 2.5 percent. The creditor probably expected to receive a real return of about 3.5 percent, and the debtor expected to pay this real return. In fact, over the life of the mortgage, the inflation rate averaged 5 percent, so the *ex post* real return was only 1 percent. This unanticipated inflation benefited the debtor at the expense of the creditor.

Unanticipated inflation also hurts people on fixed pensions. Workers and firms often agree on a fixed nominal pension when the worker retires (or even earlier). Because the pension is deferred earnings, the worker is essentially providing the firm a loan: the worker provides labor services to the firm while young but does not get fully paid until old age. Like any creditor, the worker is hurt when inflation is higher than anticipated. Like any debtor, the firm is hurt when inflation is lower than anticipated.

These situations provide a clear argument against variable inflation. The more variable the rate of inflation, the greater the uncertainty that both debtors and creditors face. Because most people are *risk averse*—they dislike uncertainty—the unpredictability caused by highly variable inflation hurts almost everyone.

Given the effects of uncertain inflation, it is puzzling that nominal contracts are so common. One might expect debtors and creditors to protect themselves from this uncertainty by writing contracts in real terms—that is, by indexing to some measure of the price level. In economies with high and variable inflation, indexation is often widespread; sometimes this indexation takes the form of writing contracts using a more stable foreign currency. In economies with moderate inflation, such as the United States, indexation is rare. Yet even in the United States, some long-term obligations are indexed. For example, Social Security benefits for the elderly are adjusted annually in response to changes in the consumer price index. And in 1997, the U.S. federal government issued inflation-indexed bonds for the first time.

Finally, in thinking about the costs of inflation, we should note a widely documented but little understood fact: high inflation is variable inflation. That is, countries with high average inflation also tend to have inflation rates that change greatly from year to year. The implication is that if a country decides to pursue a high-inflation monetary policy, it will likely have to accept highly variable inflation as well. As we have discussed, highly variable inflation increases uncertainty for both creditors and debtors by subjecting them to arbitrary and potentially large redistributions of wealth.

CASE STUDY

The Free Silver Movement, the Election of 1896, and the Wizard of Oz

The redistributions of wealth caused by unexpected changes in the price level are often a source of political turmoil, as evidenced by the Free Silver movement in the late nineteenth century. From 1880 to 1896, the price level in the United States fell 23 percent. This deflation was good for creditors, primarily the bankers of the Northeast, but it was bad for debtors, primarily the farmers of the South and West. One proposed solution to this problem was to replace the gold standard with a bimetallic standard, under which both gold and silver could be minted into coin. The move to a bimetallic standard would increase the money supply and stop the deflation.

The silver issue dominated the presidential election of 1896. William McKinley, the Republican nominee, campaigned on a platform of preserving the gold standard. William Jennings Bryan, the Democratic nominee, supported the bimetallic standard. In a famous speech, Bryan proclaimed, “You shall not press down upon the brow of labor this crown of thorns, you shall not crucify mankind upon a cross of gold.” Not surprisingly, McKinley was the candidate of the conservative eastern establishment, whereas Bryan was the candidate of the southern and western populists.

This debate over silver found its most memorable expression in a children’s book, *The Wizard of Oz*. Written by a Midwestern journalist, L. Frank Baum, just after the 1896 election, it tells the story of Dorothy, a girl lost in a strange land far from her home in Kansas. Dorothy (representing traditional American values) makes three friends: a scarecrow (the farmer), a tin woodman (the industrial worker), and a lion whose roar exceeds his might (William Jennings Bryan). Together, they make their way along a perilous yellow brick road (the gold standard), hoping to find the Wizard who will help Dorothy return home. Eventually they arrive in Oz (Washington), where everyone sees the world through green glasses (money). The Wizard (William McKinley) tries to be all things to all people but turns out to be a fraud. Dorothy’s problem is solved only when she learns about the magical power of her silver slippers.[7]

The Republicans won the election of 1896, and the United States stayed on a gold standard, but the Free Silver advocates got the inflation that they wanted. Around the time of the election, gold was discovered in Alaska, Australia, and South Africa. In addition, gold refiners devised the cyanide process, which facilitated the extraction of gold from ore. These developments led to increases in the money supply and in prices. From 1896 to 1910 the price level rose 35 percent. ■

One Benefit of Inflation

So far, we have discussed the many costs of inflation. These costs lead many economists to conclude that monetary policymakers should aim for zero inflation. Yet there is another side to the story. Some economists believe that a little bit of inflation—say, 2 or 3 percent per year—can be a good thing.

The argument for moderate inflation starts with the observation that cuts in nominal wages are rare: firms are reluctant to cut their workers' nominal wages, and workers are reluctant to accept such cuts. A 2 percent wage cut in a zero-inflation world is, in real terms, the same as a 3 percent raise with 5 percent inflation, but workers do not always see it that way. The 2 percent wage cut may seem like an insult, whereas the 3 percent raise is, after all, still a raise. Empirical studies confirm that nominal wages rarely fall.

This finding suggests that some inflation may make labor markets work better. The supply and demand for different kinds of labor are always changing. Sometimes an increase in supply or decrease in demand leads to a fall in the equilibrium real wage for a group of workers. If nominal wages can't be cut, then the only way to cut real wages is to allow inflation to do the job. Without inflation, the real wage will be stuck above the equilibrium level, resulting in higher unemployment.

For this reason, some economists argue that inflation "greases the wheels" of labor markets. Only a little inflation is needed: an inflation rate of 2 percent lets real wages fall by 2 percent per year, or about 20 percent per decade, without cuts in nominal wages. Such automatic reductions in real wages are impossible with zero inflation.[8]

5-6 Hyperinflation

Hyperinflation is often defined as inflation that exceeds 50 percent per month, which is just over 1 percent per day. Compounded over many months, this rate of inflation leads to very large increases in the price level. An inflation rate of 50 percent per month implies a more than 100-fold increase in the price level over a year and a more than 2-million-fold increase over three years. Here we consider the costs and causes of such extreme inflation.

The Costs of Hyperinflation

Although economists debate whether the costs of moderate inflation are large or small, no one doubts that hyperinflation extracts a high toll on society. The costs are qualitatively the same as those we discussed earlier. When inflation reaches extreme levels, however, these costs are more apparent because they are so severe.

The shoeleather costs from reduced money holding, for instance, are serious under hyperinflation. Business executives devote much time and energy to cash management when cash loses its value quickly. By diverting this time and energy from more socially valuable activities, such as production and investment decisions, hyperinflation makes the economy run less efficiently.

Menu costs also become larger under hyperinflation. Firms must change prices so often that normal business practices, such as printing and distributing catalogs with fixed prices, become impossible. In one restaurant during the German hyperinflation of the 1920s, a waiter would stand up on a table every 30 minutes to call out the new prices.

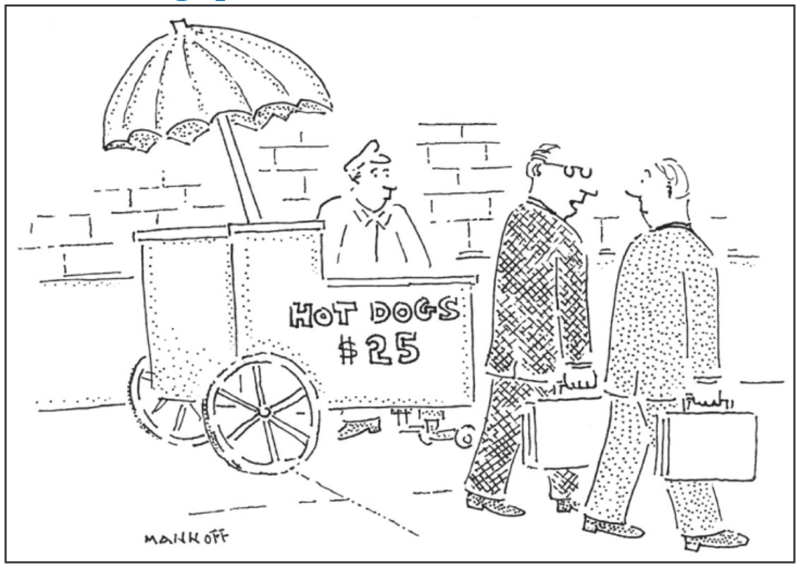
Similarly, relative prices do not do a good job of reflecting true scarcity during hyperinflations. When prices change frequently by large amounts, it is hard for customers to shop around for the best price. Highly volatile and rapidly rising prices can alter behavior in many ways. According to one report, when patrons entered a pub during the German hyperinflation, they would often buy two pitchers of beer. Although the second pitcher would lose value by getting warm over time, it would lose value less rapidly than the money left sitting in the patron’s wallet.

Tax systems are also distorted by hyperinflation—but in ways that are different from the distortions of moderate inflation. In most tax systems, there is a delay between the time a tax is levied and the time it is paid to the government. In the United States, for example, taxpayers are required to make estimated income tax payments every three months. This short delay does not matter much under low inflation. By contrast, during hyperinflation, even a short delay greatly reduces real tax revenue. By the time the government gets the money it is due, the money has fallen in value. As a result, once hyperinflations start, the real tax revenue of the government often falls substantially.

Finally, no one should underestimate the sheer inconvenience of living with hyperinflation. When carrying money to the grocery store is as burdensome as carrying the groceries back home, the monetary system is not doing its best to facilitate exchange. The government tries to overcome this problem by adding more and more zeros to the paper currency, but often it cannot keep up with the exploding price level.

Eventually, these costs of hyperinflation become intolerable. Over time, money loses its role as a store of value, unit of account, and medium of exchange. Barter becomes more common. And more stable unofficial monies—cigarettes or the U.S. dollar—start to replace the official money.

The Causes of Hyperinflation



“I told you the Fed should have tightened”
Bob Mankoff/Cartoonstock.com

Why do hyperinflations start, and how do they end? This question can be answered at different levels.

The most obvious answer is that hyperinflations are due to excessive growth in the supply of money. When the central bank prints money, the price level rises. When it prints money rapidly enough, the result is hyperinflation. To stop the hyperinflation, the central bank must reduce the rate of money growth.

This answer is incomplete, however, for it leaves open the question of why central banks in hyperinflating economies choose to print so much money. To address this deeper question, we must turn our attention from monetary to fiscal policy. Most hyperinflations begin when the government has inadequate tax revenue to pay for its spending. Although the government might prefer to finance this budget deficit by issuing debt, it may find itself unable to borrow, perhaps because lenders view the government as a bad credit risk. To cover the deficit, the government turns to the only mechanism at its disposal—the printing press. The result is rapid money growth and hyperinflation.

Once the hyperinflation is under way, the fiscal problems become even more severe. Because of the delay in collecting tax payments, real tax revenue falls as inflation rises. Thus, the government's need to rely on seigniorage is self-reinforcing. Rapid money creation leads to hyperinflation, which leads to a larger budget deficit, which leads to even more rapid money creation.

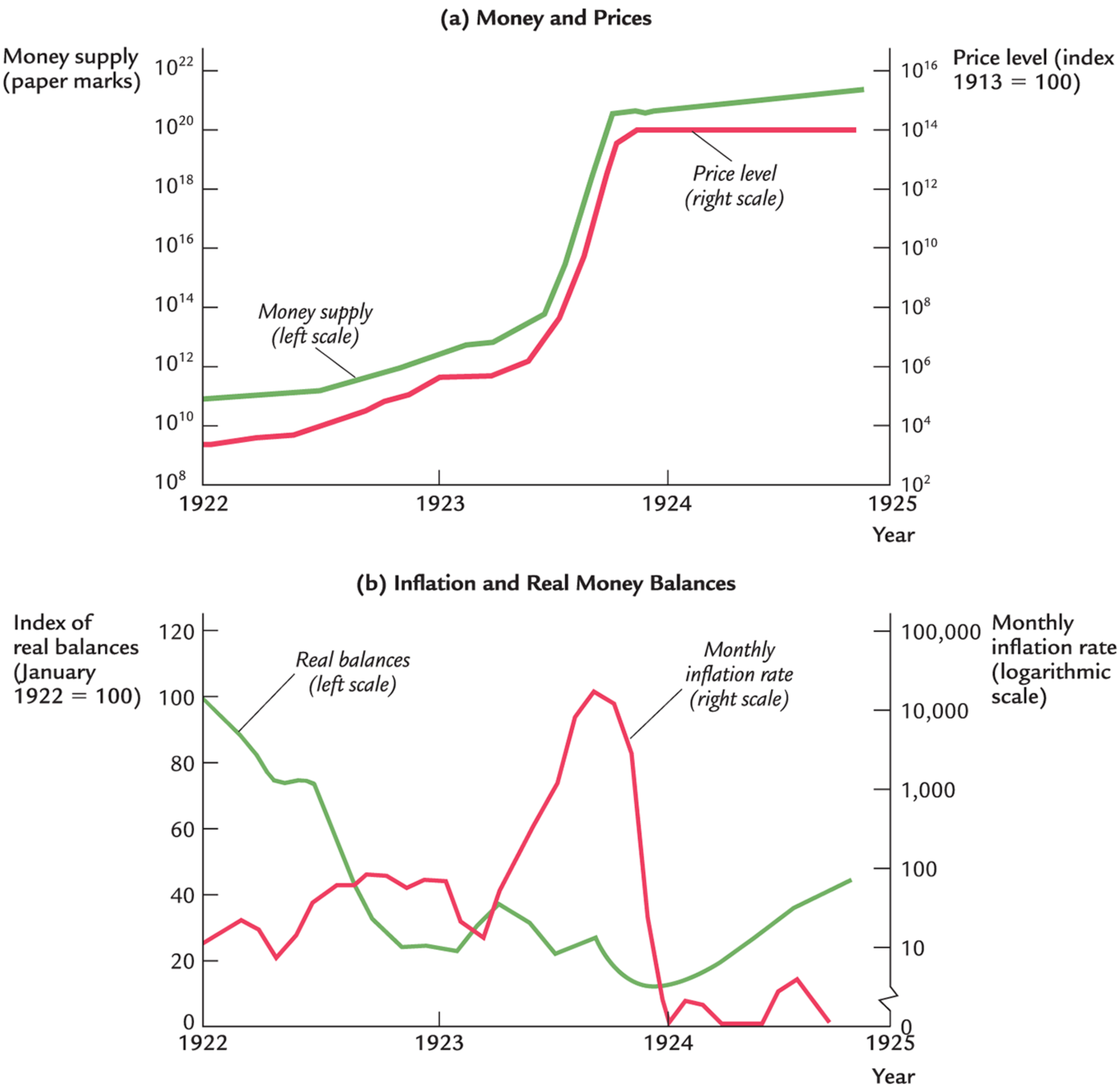
The ends of hyperinflations almost always coincide with fiscal reforms. Once the magnitude of the problem becomes apparent, the government musters the political will to reduce government spending and increase taxes. These fiscal reforms reduce the need for seigniorage, which allows a reduction in money growth. Hence, even if "inflation is always and everywhere a monetary phenomenon," as Milton Friedman suggested, the end of hyperinflation is often a fiscal phenomenon as well.[9]

CASE STUDY

Hyperinflation in Interwar Germany

After World War I, Germany experienced one of history's most spectacular examples of hyperinflation. At the war's end, the Allies demanded that Germany pay substantial reparations. These payments led to fiscal deficits in Germany, which the German government eventually financed by printing large quantities of money.

Panel (a) of Figure 5-6 shows the quantity of money and the general price level in Germany from January 1922 to December 1924. During this period both money and prices rose at an amazing rate. For example, the price of a daily newspaper rose from 0.30 mark in January 1921 to 1 mark in May 1922, to 8 marks in October 1922, to 100 marks in February 1923, and to 1,000 marks in September 1923. Then, in the fall of 1923, prices took off: the newspaper sold for 2,000 marks on October 1, 20,000 marks on October 15, 1 million marks on October 29, 15 million marks on November 9, and 70 million marks on November 17. In December 1923 the money supply and prices abruptly stabilized.[10]



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 5-6 Money and Prices in Interwar Germany Panel (a) shows the money supply and the price level in Germany from January 1922 to December 1924. The immense increases in the money supply and the price level provide a dramatic illustration of the effects of printing large amounts of money. Panel (b) shows inflation and real money balances. As inflation rose, real money balances fell. When the inflation ended at the end of 1923, real money balances rose.

Data from: Adapted from Thomas J. Sargent, “The End of Four Big Inflations,” in Inflation, edited by Robert Hall (Chicago: University of Chicago Press, 1983), 41–98.

Just as fiscal problems caused the German hyperinflation, a fiscal reform ended it. At the end of 1923, the number of government employees was cut by one-third, and the reparations payments were temporarily suspended and eventually reduced. At the same time, a new central bank, the Rentenbank, replaced the old central bank, the Reichsbank. The Rentenbank was committed to not financing the government by printing money.

According to our theoretical analysis of money demand, an end to a hyperinflation should lead to an increase in real money balances as the cost of holding money falls. Panel (b) of Figure 5-6 shows that real money balances in Germany did fall as inflation increased and then increased again as inflation fell. Yet the increase in real money balances was not immediate. Perhaps the adjustment of real money balances to the cost of holding money is a gradual process. Or perhaps it took time for people in Germany to believe that the inflation had ended, so that expected inflation fell more gradually than actual inflation. ■

CASE STUDY

Hyperinflation in Zimbabwe

In 1980, after years of colonial rule, the old British colony of Rhodesia became the new African nation of Zimbabwe. A new currency, the Zimbabwe dollar, was introduced to replace the Rhodesian dollar. For the first decade, inflation in the new nation was modest—about 10 to 20 percent per year. That, however, would soon change.

The hero of the Zimbabwe independence movement was Robert Mugabe. In general elections in 1980, he became the nation's first prime minister and later, after a government reorganization, its president. Over the years, he continued to get reelected. In his 2008 reelection, however, there were widespread claims of electoral fraud and threats against voters who supported rival candidates. At the age of 84, Mugabe was no longer as popular as he once was, but he gave no sign of any willingness to relinquish power.

Throughout his tenure, Mugabe's economic philosophy was Marxist, and one of his goals was to redistribute wealth. In the 1990s his government instituted a series of land reforms with the ostensible purpose of redistributing land from the white minority who ruled Zimbabwe during the colonial era toward the historically disenfranchised black population. One result of these reforms was widespread corruption. Many abandoned and expropriated white farms ended up in the hands of cabinet ministers and senior government officials. Another result was a substantial decline in farm output. Productivity fell as many of the experienced white farmers fled the country.

The decline in the economy's output led to a fall in the government's tax revenue. The government responded to this revenue shortfall by printing money to pay the salaries of government employees. As textbook economic theory predicts, the monetary expansion led to higher inflation.

Mugabe tried to deal with inflation by imposing price controls. Once again, the result was predictable: a shortage of many goods and the growth of an underground economy where price controls and tax collection were evaded. The government's tax revenue declined further, inducing even more monetary expansion and yet higher inflation. In July 2008, the officially reported inflation rate was 231 million percent (about 4 percent per day), though some observers put it higher. Official inflation data were soon suspended, but unofficial reports indicate that inflation continued to accelerate and, by the end of 2008, was completely out of control.

The repercussions of the hyperinflation were widespread. In an article in the *Washington Post*, one Zimbabwean citizen described the situation as follows: "If you don't get a bill collected in 48 hours, it isn't worth collecting, because it is worthless. Whenever we get money, we must immediately spend it, just go and buy what we can. Our pension was destroyed ages ago. None of us have any savings left."

The Zimbabwe hyperinflation finally ended in March 2009, when the government abandoned its own money.

The U.S. dollar became the nation’s official currency. Inflation quickly stabilized and remained low in the years that followed. Zimbabwe still had problems, but at least hyperinflation was not among them. ■

5-7 Conclusion: The Classical Dichotomy

Over the course of this and the previous chapter, we have studied the meaning of money and the impact of the money supply on inflation and various other variables. This analysis builds on our model of national income in Chapter 3. Let's now step back and examine a key assumption that has been implicit in our discussion.

In Chapter 3, we explained many macroeconomic variables. Some of these variables were *quantities*, such as real GDP and the capital stock; others were *relative prices*, such as the real wage and the real interest rate. But all of these variables had one thing in common—they measured a physical (rather than a monetary) quantity. Real GDP is the quantity of goods and services produced in a given year, and the capital stock is the quantity of machines and structures available at a given time. The real wage is the quantity of output a worker earns for each hour of work, and the real interest rate is the quantity of output a person earns in the future by lending one unit of output today. All variables measured in physical units, such as quantities and relative prices, are called **real variables**.

In this chapter we examined **nominal variables**—variables expressed in terms of money. The economy has many nominal variables, such as the price level, the inflation rate, and the dollar wage a person earns.

At first, it may seem surprising that we could explain real variables without introducing nominal variables or the existence of money. In Chapter 3 we studied the level and allocation of the economy's output without mentioning the price level or the rate of inflation. Our theory of the labor market explained the real wage without explaining the nominal wage.

Economists call this theoretical separation of real and nominal variables the **classical dichotomy**. It is the hallmark of classical macroeconomic theory. The classical dichotomy is an important insight because it simplifies economic theory. It allows us to examine real variables, as we have done, while ignoring nominal variables. The classical dichotomy arises because, in classical economic theory, changes in the money supply do not influence real variables. This irrelevance of money in the determination of real variables is called **monetary neutrality**. For many purposes—in particular for studying long-run issues—monetary neutrality is approximately correct.

Yet monetary neutrality does not fully describe the world in which we live. Beginning in Chapter 10, we discuss departures from the classical model and monetary neutrality. These departures are crucial for understanding many macroeconomic phenomena, such as short-run economic fluctuations.

CHAPTER 6

The Open Economy

> No nation was ever ruined by trade.
>
> —Benjamin Franklin

Even if you never leave your hometown, you are a participant in the global economy. When you go to the grocery store, you might choose between apples grown locally and grapes grown in Chile. When you make a deposit into your local bank, the bank might lend those funds to your next-door neighbor or to a Japanese company building a factory outside Tokyo. Because economies around the world are integrated with one another, consumers have more goods and services from which to choose, and savers have more opportunities to invest their wealth.

In previous chapters we simplified the analysis by assuming a closed economy. Yet most actual economies are open: they export goods and services abroad, they import goods and services from abroad, and they borrow and lend in world financial markets. Figure 6-1 gives some sense of the importance of these international interactions by showing imports and exports as a percentage of GDP for ten major countries. As the figure shows, exports from the United States are about 13 percent of GDP, and imports are about 15 percent. Trade is even more important for many other countries—imports and exports are about 20 percent of GDP in China, about 33 percent in Canada, and almost 50 percent in Germany. In these countries, international trade is central to analyzing economic developments and formulating economic policies.

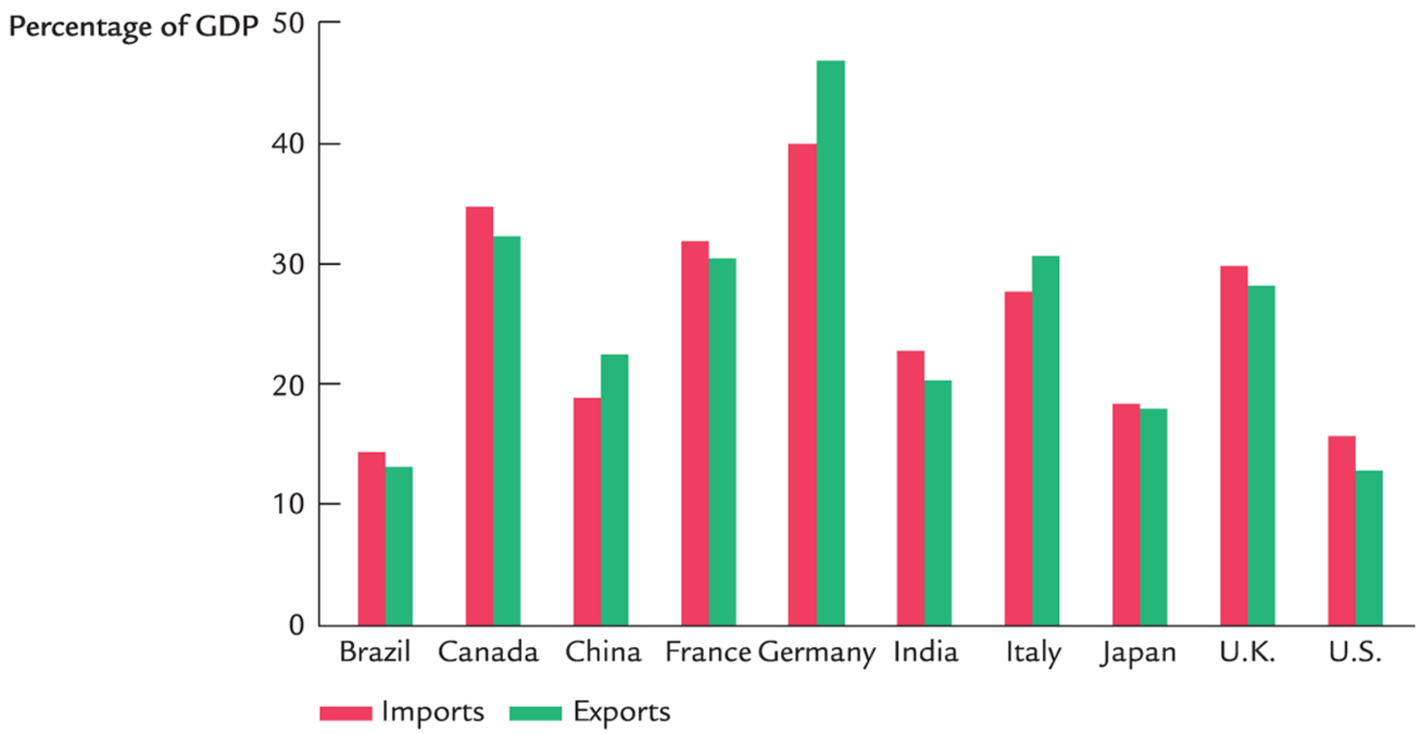


FIGURE 6-1 Imports and Exports as a Percentage of Output: 2015 While international trade is important for the United States, it is even more vital for other countries.

Data from: World Bank.

This chapter begins our study of open-economy macroeconomics. We begin in Section 6-1 with questions of measurement. To understand how an open economy works, we must understand the key macroeconomic variables that measure the interactions among countries. Accounting identities reveal a key insight: the flow of goods and services across national borders is always matched by an equivalent flow of funds to finance capital accumulation.

In Section 6-2 we examine the determinants of these international flows. We develop a model of the small open economy that corresponds to our model of the closed economy in Chapter 3. The model shows the factors that determine whether a country is a borrower or a lender in world markets and how policies at home and abroad affect the flows of capital and goods.

In Section 6-3 we extend the model to discuss the prices at which a country makes exchanges in world markets. We examine what determines the price of domestic goods relative to foreign goods. We also examine what determines the rate at which the domestic currency trades for foreign currencies. Our model shows how protectionist trade policies—policies designed to protect domestic industries from foreign competition—influence the amount of international trade and the exchange rate.

6-1 The International Flows of Capital and Goods

The key macroeconomic difference between open and closed economies is that, in an open economy, a country's spending in any given year need not equal its output of goods and services. A country can spend more than it produces by borrowing from abroad, or it can spend less than it produces and lend the difference to foreigners. To understand this more fully, let's take another look at national income accounting, which we first discussed in Chapter 2.

The Role of Net Exports

Consider the expenditure on an economy's output of goods and services, again denoted as *Y*. In a closed economy, all output is sold domestically, and expenditure is divided into three components: consumption *C*, investment *I*, and government purchases *G*. In an open economy, some output is sold domestically and some is exported to be sold abroad. In addition, some of the goods and services included in consumption, investment, and government purchases are produced abroad and imported. We can thus write the national income accounts identity as

$$Y = C + I + G + X - IM$$

where *X* represents exports and *IM* represents imports. Because spending on imports is included in domestic spending $(C + I + G)$, and because goods and services imported from abroad are not part of a country's output, this equation subtracts spending on imports.

Defining **net exports** to be exports minus imports $(NX = X - IM)$, we can write the identity as

$$Y = C + I + G + NX.$$

This equation states that expenditure on domestic output is the sum of consumption, investment, government purchases, and net exports. This form of the national income accounts identity should be familiar from Chapter 2.

The national income accounts identity shows how domestic output, domestic spending, and net exports are related. In particular,

$$\begin{array}{ccccc} NX & = & Y & - & (C+I+G) \\ \text{Net Exports} & = & \text{Output} & - & \text{Domestic Spending}. \end{array}$$

This equation shows that in an open economy, domestic spending need not equal the output of goods and services. *If a country's output exceeds its domestic spending, it exports the difference: net exports are positive. If a country's output falls short of its domestic spending, it imports the difference: net exports are negative.*

International Capital Flows and the Trade Balance

In an open economy, as in the closed economy we discussed in Chapter 3, financial markets and goods markets are closely related. To see the relationship, we must rewrite the national income accounts identity in terms of saving and investment. Begin with the identity

$$Y = C + I + G + NX.$$

Subtract *C* and *G* from both sides to obtain

$$Y - C - G = I + NX.$$

Recall from Chapter 3 that $Y - C - G$ is national saving *S*, which equals the sum of private saving, $Y - T - C$, and public saving, $T - G$, where *T* stands for taxes. Therefore,

$$S = I + NX.$$

Subtracting *I* from both sides of the equation, we can write the national income accounts identity as

$$S - I = NX.$$

This form of the national income accounts identity shows that an economy's net exports must always equal the difference between its saving and its investment.

Let's look more closely at each part of this identity. The right-hand side, *NX*, is net exports of goods and services. Another name for net exports is the **trade balance**, because it tells us how a country's trade in goods and services departs from the benchmark of equal imports and exports.

The left-hand side of the identity is the difference between domestic saving and domestic investment, $S - I$, which we'll call **net capital outflow**. (It's sometimes called *net foreign investment.*) Net capital outflow equals the amount that domestic residents are lending abroad minus the amount that foreigners are lending to us. If net capital outflow is positive, the economy's saving exceeds its investment, and it is lending the excess to foreigners. If the net capital outflow is negative, the economy is experiencing a capital inflow: investment exceeds saving, and the economy is financing this extra investment by borrowing from abroad. Thus, net capital outflow reflects the international flow of funds to finance capital accumulation.

The national income accounts identity shows that net capital outflow always equals the trade balance. That is,

$$\begin{array}{ccc} \text{Net Capital Outflow} & = & \text{Trade Balance} \\ S - I & = & NX. \end{array}$$

If $S - I$ and *NX* are positive, a country has a **trade surplus**. In this case, it is a net lender in world financial markets, and it exports more than it imports. If $S - I$ and *NX* are negative, a country has a **trade deficit**. In this case, it is a net borrower in world financial markets, and it imports more than it exports. If $S - I$ and *NX* are exactly zero, a country is said to have **balanced trade** because its imports and exports are equal in value.

The national income accounts identity shows that the international flow of funds to finance capital accumulation and the international flow of goods and services are two sides of the same coin. Suppose that, in the nation of Westeros, saving exceeds investment. In this case, the surplus saving of Westeros is used to make loans to foreigners. Foreigners require these loans because Westeros is providing them with more goods and services than they are providing Westeros. That is, Westeros is running a trade surplus. Conversely, suppose that, in the nation of Essos, investment exceeds saving. Then the extra investment in Essos must be financed by borrowing from abroad. These foreign loans enable Essos to import more goods and services than it exports. That is, Essos is running a trade deficit. Table 6-1 summarizes these lessons.

TABLE 6-1 International Flows of Goods and Capital: Summary

This table shows the three outcomes that an open economy can experience.

Trade Surplus	Balanced Trade	Trade Deficit
Exports > Imports	Exports = Imports	Exports < Imports
Net Exports>0	Net Exports=0	Net Exports<0
Y>C+I+G	Y=C+I+G	Y<C+I+G
Saving > Investment	Saving = Investment	Saving < Investment
Net Capital Outflow>0	Net Capital Outflow=0	Net Capital Outflow<0

The international flow of capital can take many forms. It is easiest to assume—as we have done so far—that when a country runs a trade deficit, foreigners lend to it. This happens, for example, when the Chinese buy the debt issued by U.S. corporations or the U.S. government. But the flow of capital can also take the form of foreigners buying domestic assets, such as when a German buys stock from an American on the New York Stock Exchange. Whether foreigners buy domestically issued debt or domestically owned assets, they obtain a claim to the future returns to domestic capital. In both cases, foreigners end up owning some of the domestic capital stock.

International Flows of Goods and Capital: An Example

The equality of net exports and net capital outflow is an identity: it must hold because of how the variables are defined and the numbers are added up. But it is easy to miss the intuition behind this important relationship. The best way to understand it is to consider an example.

Imagine that Bill Gates sells a copy of the Windows operating system to a Japanese consumer for 10,000 yen. Because Mr. Gates is a U.S. resident, the sale represents an export of the United States. Other things equal, U.S. net exports rise. What else happens to make the identity hold? It depends on what Mr. Gates does with the 10,000 yen.

Suppose Mr. Gates decides to stuff the 10,000 yen in his mattress. In this case, Mr. Gates has allocated some of his saving to an investment in the Japanese economy (in the form of the Japanese currency) rather than to an investment in the U.S. economy. Thus, U.S. saving exceeds U.S. investment. The rise in U.S. net exports is matched by a rise in the U.S. net capital outflow.

If Mr. Gates wants to invest in Japan, however, he is unlikely to make currency his asset of choice. He might use the 10,000 yen to buy some stock in, say, the Japanese firm Sony, or he might buy a bond issued by the Japanese government. In either case, some of U.S. saving flows abroad. Once again, the U.S. net capital outflow exactly balances U.S. net exports.

The opposite situation occurs in Japan. When the Japanese consumer buys a copy of the Windows operating system, Japan's purchases of goods and services (C+I+G)$(C+I+G)$ rise, but there is no change in what Japan has produced (*Y*). Japan's imports increase, and its net exports decrease. In addition, the transaction reduces Japan's saving (S=Y−C−G)$(S=Y-C-G)$ for a given level of investment (*I*). While the United States experiences a net capital outflow, Japan experiences a net capital inflow.

Now let's change the example. Suppose that instead of investing his 10,000 yen in a Japanese asset, Mr. Gates uses it to buy something made in Japan, such as a supersize box of Pokémon cards. In this case, imports into the United States rise. Together, the Windows export and the Pokémon import represent balanced trade between Japan and the United States. Because exports and imports rise equally, net exports and net capital outflow are both unchanged.

A final possibility is that Mr. Gates exchanges his 10,000 yen for U.S. dollars at a local bank. But this doesn't change the situation: the bank now must do something with the 10,000 yen. It can buy Japanese assets (a U.S. net capital outflow); it can buy a Japanese good (a U.S. import); or it can sell the yen to another American who wants to make such a transaction. If you follow the money, you can see that, in the end, U.S. net exports must equal U.S. net capital outflow.

The Irrelevance of Bilateral Trade Balances

The trade balance we have been discussing measures the difference between a nation's exports and its imports with the rest of the world. Sometimes you might hear a media report on a nation's trade balance with another nation. This is called a *bilateral* trade balance. For example, the U.S. bilateral trade balance with China equals exports that the United States sells to China minus imports that the United States buys from China.

The overall trade balance is, as we have seen, inextricably linked to a nation's saving and investment. That is not true of a bilateral trade balance. Indeed, a nation can have large trade deficits and surpluses with specific trading partners while having balanced trade overall.

For example, suppose the world has three countries: the United States, China, and Australia. The United States sells $100 billion in machine tools to Australia, Australia sells $100 billion in wheat to China, and China sells $100 billion in toys to the United States. In this case, the United States has a bilateral trade deficit with China, China has a bilateral trade deficit with Australia, and Australia has a bilateral trade deficit with the United States. But each of the three nations has balanced trade overall because it has exported and imported

$100 billion in goods.

Bilateral trade deficits receive more attention in the political arena than they deserve. This is in part because international relations are conducted country to country, so politicians and diplomats are naturally drawn to statistics measuring country-to-country economic transactions. Most economists, however, believe that bilateral trade balances are not very meaningful. From a macroeconomic standpoint, it is a nation's trade balance with all foreign nations put together that matters.

The same lesson applies to individuals as it does to nations. Your own personal trade balance is the difference between your income and your spending, and you may be concerned if these two variables are out of line. But you should not be concerned with the difference between your income and spending with a particular person or firm. Economist Robert Solow once explained the irrelevance of bilateral trade balances as follows: "I have a chronic deficit with my barber, who doesn't buy a darned thing from me." But that doesn't stop Mr. Solow from living within his means—or getting a haircut when he needs it.

6-2 Saving and Investment in a Small Open Economy

So far in our discussion of the international flows of goods and capital, we have rearranged accounting identities. That is, we have defined some of the variables that measure transactions in an open economy, and we have shown the links among these variables that follow from their definitions. Our next step is to develop a model to explain these variables. The model will answer questions such as how the trade balance responds to changes in policy.

Capital Mobility and the World Interest Rate

In a moment we present a model of the international flows of capital and goods. Because the trade balance equals the net capital outflow, which in turn equals saving minus investment, our model focuses on saving and investment. To develop this model, we use some elements that should be familiar from Chapter 3, but unlike with the Chapter 3 model, we do not assume that the real interest rate equilibrates saving and investment. Instead, we allow the economy to run a trade deficit and borrow from other countries or to run a trade surplus and lend to other countries.

If the real interest rate does not adjust to equilibrate saving and investment in this model, what *does* determine the real interest rate? We answer this question here by considering the simple case of a **small open economy** with perfect capital mobility. By "small" we mean that this economy is a small part of the world market and thus, by itself, has only a negligible effect on the world interest rate. By "perfect capital mobility" we mean that residents of the country have full access to world financial markets. In particular, the government does not impede international borrowing or lending.

Because of this assumption of perfect capital mobility, the interest rate in our small open economy *r*, must equal the **world interest rate** r*,r^*, the real interest rate prevailing in world financial markets:

r=r*.$$r = r^*.$$

Residents of the small open economy need never borrow at any interest rate above r*,r^*, because they can always get a loan at r*r^* from abroad. Similarly, residents of this economy need never lend at any interest rate below r*,r^*, because they can always earn r*r^* by lending abroad. Thus, the world interest rate determines the interest rate in a small open economy.

Let's briefly discuss what determines the world real interest rate. In a closed economy, the equilibrium of domestic saving and domestic investment determines the interest rate. Barring interplanetary trade, the world economy is a closed economy. Therefore, the equilibrium of world saving and world investment determines the world interest rate. A small open economy has a negligible effect on the world real interest rate because, being a small part of the world, it has a negligible effect on world saving and world investment. Hence, a small open economy takes the world interest rate as exogenously given.

Why Assume a Small Open Economy?

The analysis in this chapter assumes that the nation being studied is a small open economy. (Chapter 13, which examines short-run fluctuations in an open economy, takes the same approach.) This assumption raises some questions.

Q: Is the United States well described by the assumption of a small open economy?

A: No, it is not, at least not completely. The United States does borrow and lend in world financial markets, and these markets exert a strong influence over the U.S. real interest rate, but it would be an exaggeration to say that the U.S. real interest rate is determined solely by world financial markets.

Q: So why do we assume a small open economy?

A: Some nations, such as Canada and the Netherlands, are better described by the assumption of a small open economy. Yet the main reason for making this assumption is to develop understanding and intuition for the macroeconomics of open economies. Remember from Chapter 1 that models are built with simplifying assumptions. An assumption need not be realistic to be useful. Assuming a small open economy simplifies the analysis greatly and, therefore, helps clarify our thinking.

Q: Can we relax this assumption and make the model more realistic?

A: Yes, we can, and we will. The appendix to this chapter (and the appendix to Chapter 13) considers the more realistic and more complex case of a large open economy. Some instructors skip directly to this material when teaching these topics because the approach is more realistic for economies such as that of the United States. Others begin with the simplifying assumption of a small open economy.

The Model

To build the model of the small open economy, we take three assumptions from Chapter 3:

- The economy's output *Y* is fixed by its factors of production and its production function. We write this as

$$Y = \overline{Y} = F\left(\overline{K}, \overline{L}\right).$$

- Consumption *C* is positively related to disposable income $Y - T$. We write the consumption function as

$$C = C\left(Y - T\right).$$

- Investment *I* is negatively related to the real interest rate *r*. We write the investment function as

$$I = I\left(r\right).$$

These are the three key parts of our model. If you do not understand these relationships, review Chapter 3 before continuing.

We can now return to the accounting identity and write it as

$$\begin{aligned} NX &= (Y - C - G) - I \\ NX &= S - I. \end{aligned}$$

Substituting the Chapter 3 assumptions recapped above and the assumption that the interest rate equals the world interest rate, we obtain

$$\begin{aligned} NX &= \left[\overline{Y} - C\left(\overline{Y} - T\right) - G\right] - I\left(r^*\right) \\ &= \overline{S} - I\left(r^*\right). \end{aligned}$$

This equation shows that the trade balance *NX* depends on those variables that determine saving *S* and investment *I*. Because saving depends on fiscal policy (lower government purchases *G* or higher taxes *T* raise national saving) and investment depends on the world real interest rate r^* (a higher interest rate makes some investment projects unprofitable), the trade balance depends on these variables as well.

In Chapter 3, we graphed saving and investment as in Figure 6-2. In the closed economy studied in that chapter, the real interest rate adjusts to equilibrate saving and investment, so the real interest rate is found where the saving and investment curves cross. In the small open economy, however, the real interest rate equals the world real interest rate. *The trade balance is determined by the difference between saving and investment at the world interest rate*.

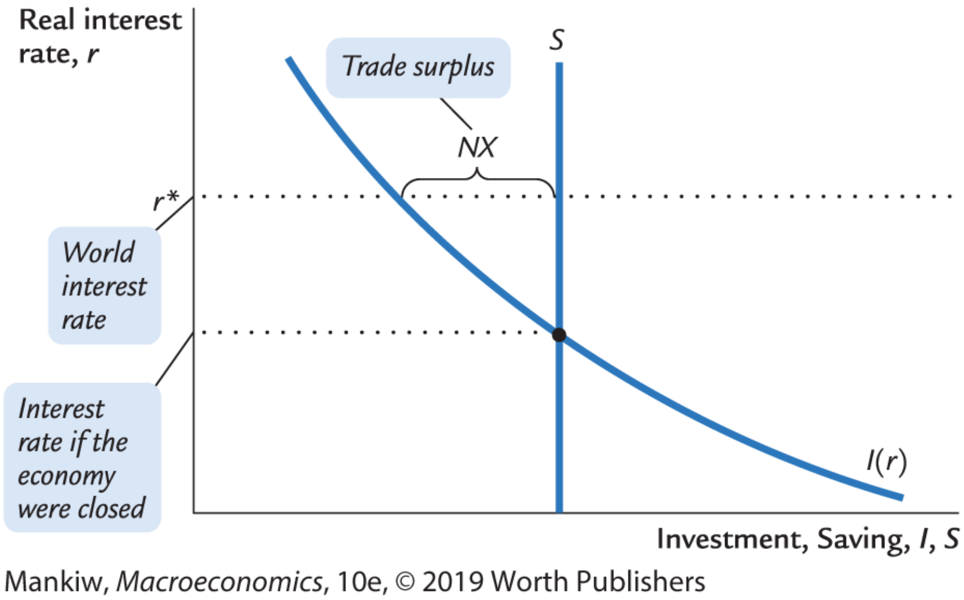


FIGURE 6-2 Saving and Investment in a Small Open Economy In a closed economy, the real interest rate adjusts to equilibrate saving and investment. In a small open economy, the interest rate is determined in world financial markets. The difference between saving and investment determines the trade balance. Here there is a trade surplus, because at the world interest rate, saving exceeds investment.

At this point, you might wonder about the mechanism that causes the trade balance to equal the net capital outflow. The determinants of the capital flows are easy to understand. When saving falls short of investment, investors borrow from abroad; when saving exceeds investment, the excess is lent to other countries. But what causes those who import and export to behave so that the international flow of goods exactly balances this international flow of capital? For now we leave this question unanswered, but we return to it in Section 6-3 when we discuss exchange rates.

How Policies Influence the Trade Balance

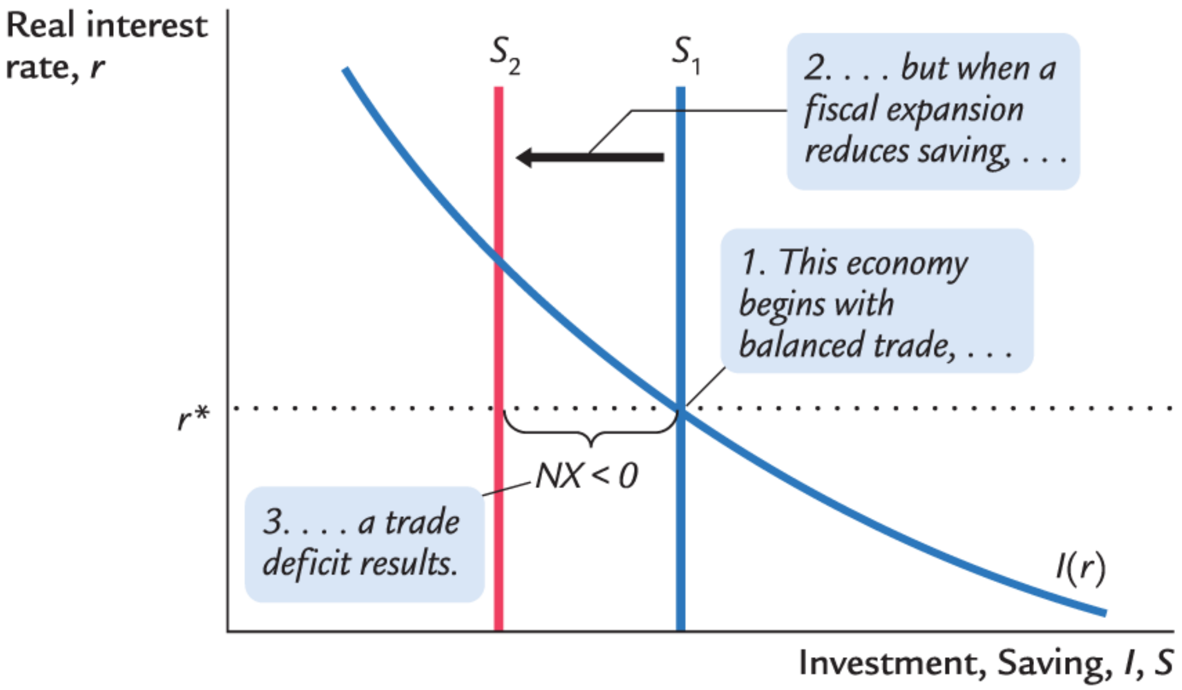
Suppose that the economy begins in a position of balanced trade. That is, at the world interest rate, investment *I* equals saving *S*, and net exports *NX* equal zero. Let’s use our model to examine the effects of government policies at home and abroad.

Fiscal Policy at Home

Consider first what happens to the small open economy if the government expands domestic spending by increasing government purchases. The increase in *G* reduces national saving, because $S=Y-C-G$. $S = Y - C - G$. With an unchanged world real interest rate, investment remains the same. Therefore, saving falls below investment, and some investment must now be financed by borrowing from abroad. Because $NX=S-I$, $NX = S - I$, the fall in *S* implies a fall in *NX*. The economy now runs a trade deficit.

The same logic applies to a decrease in taxes. A tax cut lowers *T*, raises disposable income $Y-T$, $Y - T$, stimulates consumption, and reduces national saving. (Even though some of the tax cut finds its way into private saving, public saving falls by the full amount of the tax cut; in total, saving falls.) Because $NX=S-I$, $NX = S - I$, the reduction in national saving in turn lowers *NX*.

Figure 6-3 shows these effects. A fiscal policy change that increases private consumption *C* or public consumption *G* reduces national saving $(Y-C-G)$ $(Y - C - G)$ and, therefore, shifts the vertical line that represents saving from $S1$ S_1 to $S2$. S_2. Because *NX* is the distance between the saving schedule and the investment schedule at the world interest rate, this shift reduces *NX*. *Hence, starting from balanced trade, a change in fiscal policy that reduces national saving leads to a trade deficit.*



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

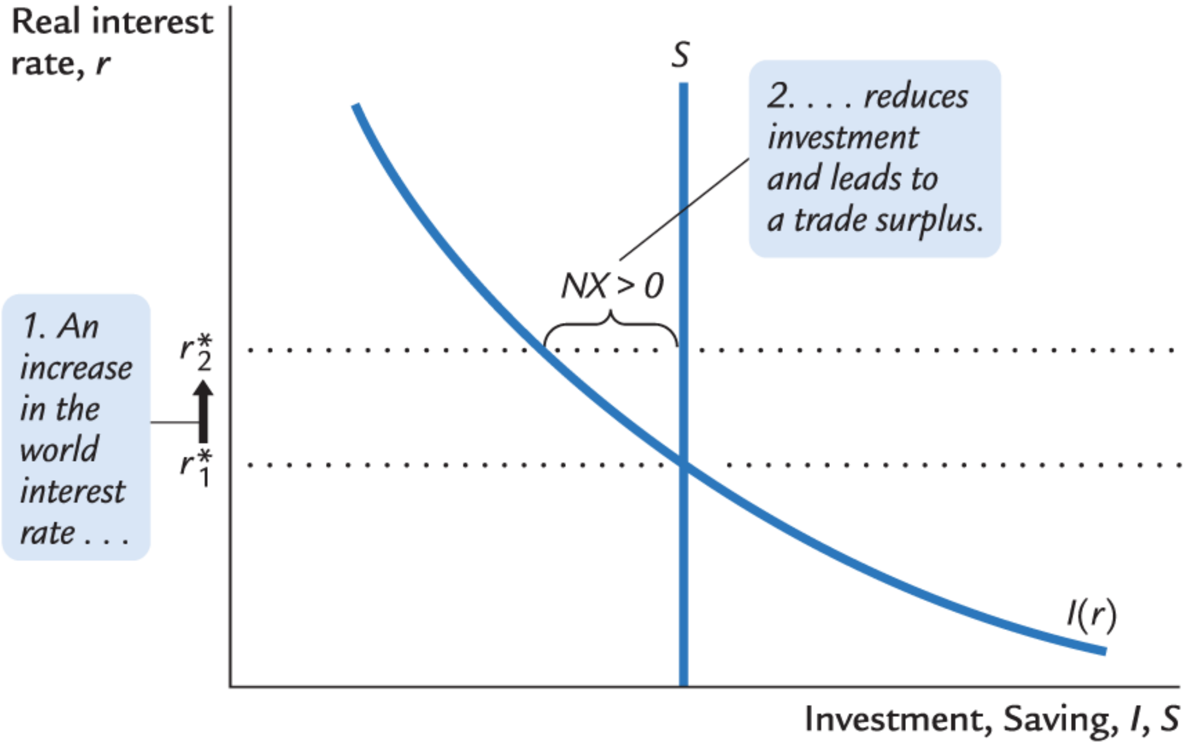
FIGURE 6-3 A Fiscal Expansion at Home in a Small Open Economy An increase in government purchases or a reduction in taxes reduces national saving and thus shifts the saving schedule to the left, from $S1$ S_1 to $S2$ S_2 The result is a trade deficit.

Fiscal Policy Abroad

Consider now what happens to a small open economy when foreign governments increase their government purchases. If these foreign countries are a small part of the world economy, then their fiscal change has a negligible impact on other countries. But if these foreign countries are a large part of the world economy, their increase in government purchases reduces world saving. The decrease in world saving causes the world interest rate to rise, just as we saw in the closed-economy model (remember, Earth is a closed economy).

The increase in the world interest rate raises the cost of borrowing and, thus, reduces investment in our small open economy. Because there has been no change in domestic saving, saving *S* now exceeds investment *I*, and some of the country's saving begins to flow abroad. Because $NX = S - I$, the reduction in *I* must also increase *NX*. Hence, reduced saving abroad leads to a trade surplus at home.

Figure 6-4 illustrates how a small open economy starting from balanced trade responds to a foreign fiscal expansion. Because the policy change occurs abroad, the domestic saving and investment schedules remain the same. The only change is an increase in the world interest rate from r_1^* to r_2^*. The trade balance is the difference between the saving and investment schedules; because saving exceeds investment at r_2^*, there is a trade surplus. *Hence, starting from balanced trade, an increase in the world interest rate due to a fiscal expansion abroad leads to a trade surplus.*

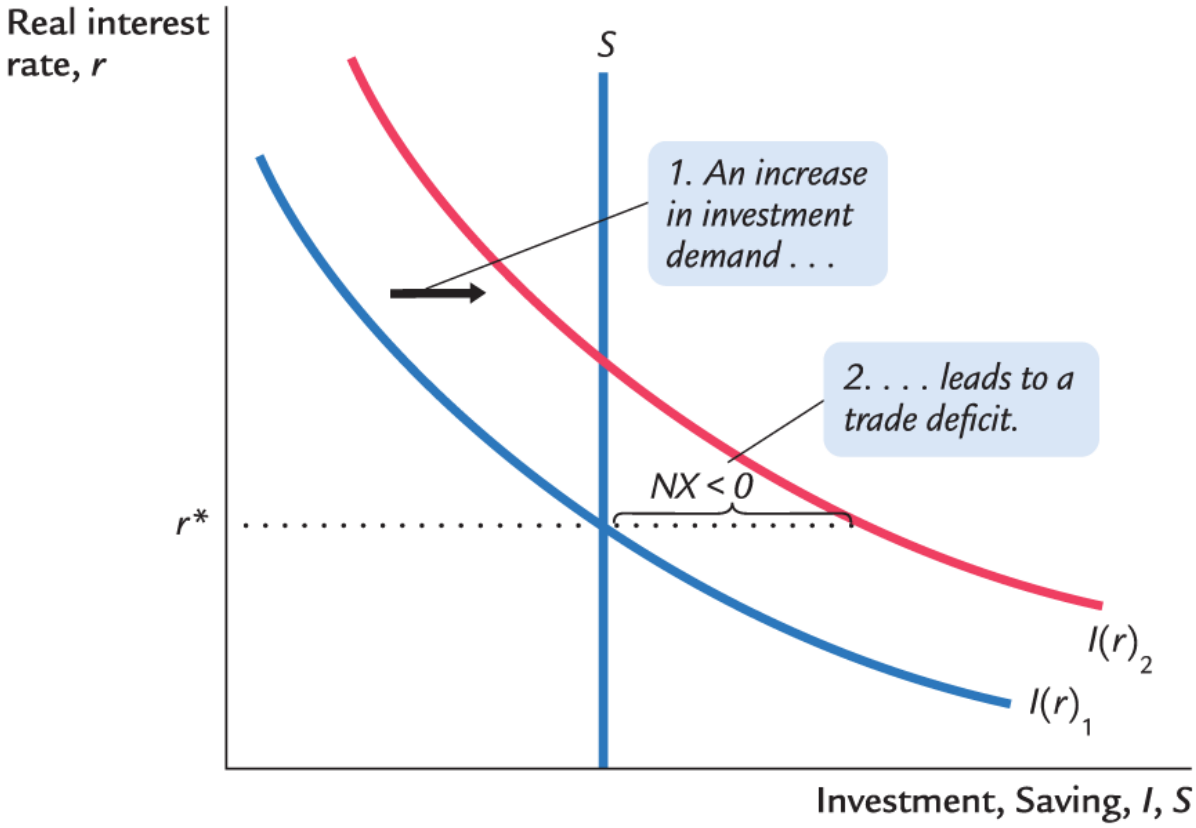


Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 6-4 A Fiscal Expansion Abroad in a Small Open Economy A fiscal expansion in a foreign economy large enough to influence world saving and investment raises the world interest rate from r_1^* to r_2^*. The higher world interest rate reduces investment in this small open economy, causing a trade surplus.

Shifts in Investment Demand

Consider what happens to a small open economy if its investment schedule shifts outward so there is greater demand for investment goods at every interest rate. This shift would occur if, for example, the government changed the tax laws to encourage investment by providing an investment tax credit. Figure 6-5 shows the impact of a shift in the investment schedule. At a given world interest rate, investment is now higher. Because saving is unchanged, some investment must now be financed by borrowing from abroad. Because capital flows into the economy to finance the increased investment, the net capital outflow is negative. Put differently, because $NX = S - I$, the increase in I implies a decrease in NX. *Hence, starting from balanced trade, an outward shift in the investment schedule causes a trade deficit.*



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 6-5 A Shift in the Investment Schedule in a Small Open Economy An outward shift in the investment schedule from $I(r)_1$ to $I(r)_2$ increases the amount of investment at the world interest rate r^*. As a result, investment now exceeds saving, which means the economy is borrowing from abroad and running a trade deficit.

Evaluating Economic Policy

Our model of the open economy shows that the flow of goods and services measured by the trade balance is inextricably connected to the international flow of funds for capital accumulation. The net capital outflow is the difference between domestic saving and domestic investment. Thus, the impact of economic policies on the trade balance can always be found by examining their impact on domestic saving and domestic investment. Policies that increase investment or decrease saving tend to cause a trade deficit, and policies that decrease investment or increase saving tend to cause a trade surplus.

Our analysis of the open economy has been positive, not normative. It has shown how various policies influence the international flows of capital and goods but not whether these policies and outcomes are desirable. Evaluating economic policies and their impact on the open economy is a frequent topic of debate among economists and policymakers.

When a country runs a trade deficit, policymakers must confront the question of whether it represents a national problem. Most economists view a trade deficit not as a problem in itself, but perhaps as a symptom of a problem. A trade deficit could reflect low saving. In a closed economy, low saving leads to low investment and a smaller future capital stock. In an open economy, low saving leads to a trade deficit and a growing foreign debt, which eventually must be repaid. In both cases, high current consumption leads to lower future consumption, implying that future generations will bear the burden of low national saving.

Yet trade deficits are not always a symptom of an economic malady. When poor rural economies develop into modern industrial economies, they sometimes finance increased investment with foreign borrowing. In these cases, trade deficits are a sign of economic development. For example, South Korea ran large trade deficits throughout the 1970s and early 1980s, and it became one of the success stories of economic growth. The lesson is that one cannot judge economic performance from the trade balance alone. Instead, one must look at the underlying causes of the international flows.

CASE STUDY

The U.S. Trade Deficit

During the 1980s, 1990s, and 2000s, the United States ran large trade deficits. Panel (a) of Figure 6-6 documents this experience by showing net exports as a percentage of GDP. The exact size of the trade deficit fluctuated over time, but it was large throughout these three decades. In 2016, the trade deficit was $452 billion, or 2.8 percent of GDP. As accounting identities require, this trade deficit had to be financed by borrowing from abroad (or, equivalently, by selling U.S. assets abroad). During this period, the United States went from being the world's largest creditor to the world's largest debtor.

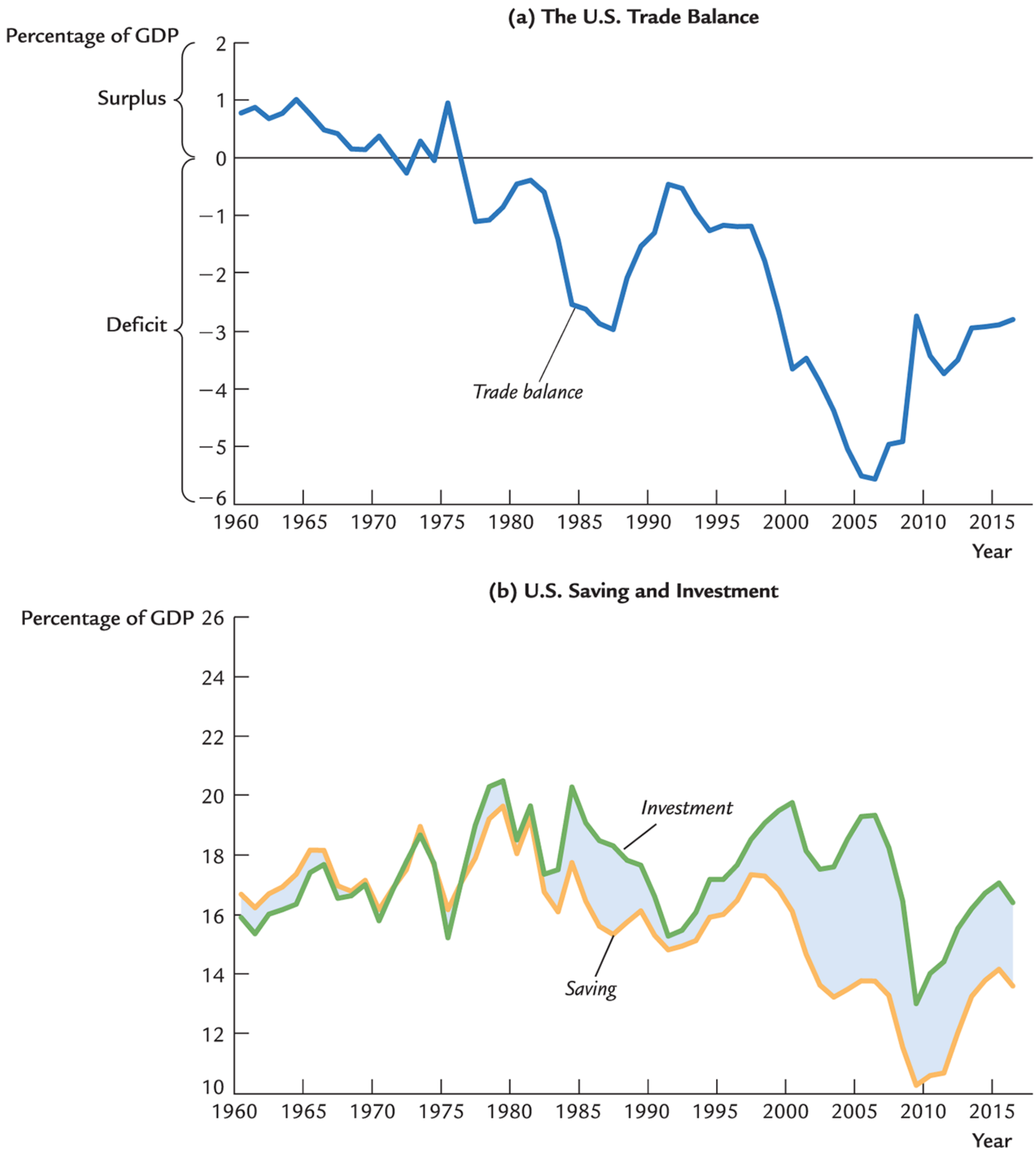


FIGURE 6-6 The Trade Balance, Saving, and Investment: The U.S. Experience Panel (a) shows the trade balance as a percentage of GDP. Positive numbers represent a surplus, and negative numbers represent a deficit. Panel (b) shows national saving and investment as a percentage of GDP from 1960 to 2016. The trade balance equals saving minus investment.

Data from: U.S. Department of Commerce.

What caused the U.S. trade deficit? There is no single explanation. But to understand some of the forces at work, it helps to look at national saving and domestic investment, as shown in panel (b) of the figure. Keep in mind that the trade deficit is the difference between saving and investment.

The start of the trade deficit coincided with a fall in national saving. This development can be explained by the expansionary fiscal policy in the 1980s. With the support of President Reagan, the U.S. Congress passed legislation in 1981 that substantially cut personal income taxes over the next three years. Because these tax cuts were not met with equal cuts in government spending, the federal budget went into deficit. These budget deficits were among the largest ever experienced in a period of peace and prosperity, and they continued long after Reagan left office. According to our model, such a policy should reduce national saving, thereby causing a trade deficit. And, in fact, that is exactly what happened. Because the government budget and trade balance went into deficit at roughly the same time, these shortfalls were called the *twin deficits*.

Things started to change in the 1990s, when the U.S. federal government got its fiscal house in order. President George H. W. Bush and President Bill Clinton both signed tax increases, while Congress kept a lid on spending. In addition to these policy changes, rapid productivity growth in the late 1990s raised incomes and, thus, further increased tax revenue. These developments moved the U.S. federal budget from deficit to surplus, which in turn caused national saving to rise.

In contrast to what our model predicts, the increase in national saving did not coincide with a shrinking trade deficit, because domestic investment rose at the same time. The likely explanation is that the boom in information technology in the 1990s caused an expansionary shift in the U.S. investment function. Even though fiscal policy was pushing the trade deficit toward surplus, the investment boom was an even stronger force pushing the trade balance toward deficit.

In the early 2000s, fiscal policy once again put downward pressure on national saving. With President George W. Bush in the White House, tax cuts were signed into law in 2001 and 2003, while the war on terror led to substantial increases in government spending. The federal government was again running budget deficits. National saving fell to historic lows, and the trade deficit reached historic highs.

A few years later, the trade deficit started to shrink somewhat, as the economy experienced a substantial decline in house prices (which led to the Great Recession, a phenomenon examined in Chapter 12). Lower house prices led to a substantial decline in residential investment. The trade deficit fell from 5.5 percent of GDP at its peak in 2006 to 3.0 percent in 2013. From 2013 to 2016, as the economy gradually recovered from the economic downturn, saving and investment both increased, with little change in the trade balance.

The history of the U.S. trade deficit shows that this statistic, by itself, does not tell us much about what is happening in the economy. We have to look deeper at saving, investment, and the policies and events that cause them (and thus the trade balance) to change over time.[1] ■

CASE STUDY

Why Doesn’t Capital Flow to Poor Countries?

The U.S. trade deficit discussed in the previous Case Study represents a flow of capital into the United States from the rest of the world. What countries were the source of these capital flows? Because the world is a closed economy, the capital must have been coming from those countries that were running trade surpluses. In 2017, this group included many nations that were poorer than the United States, such as China, Malaysia, Thailand, Estonia, and Slovenia. In these nations, saving exceeded investment in domestic capital. These countries were sending funds abroad to countries like the United States, where investment in domestic capital exceeded saving.

From one perspective, the direction of international capital flows is a paradox. Recall our discussion of production functions in Chapter 3. There, we established that an empirically realistic production function is the Cobb–Douglas form:

$$F(K, L) = A K^{\alpha} L^{1-\alpha},$$

where *K* is capital, *L* is labor, *A* is a variable representing the state of technology, and α is a parameter that determines capital's share of total income. For this production function, the marginal product of capital is

$$MPK = \alpha A (K/L)^{\alpha - 1}.$$

The marginal product of capital tells us how much extra output an extra unit of capital would produce. Because α is capital's share, it must be less than 1, so $\alpha - 1 < 0.$ This means that an increase in *K/L* decreases *MPK*. In other words, holding other variables constant, the more capital a nation has, the less valuable an extra unit of capital is. This phenomenon of diminishing marginal product says that capital should be more valuable where capital is scarce.

This prediction, however, seems at odds with the international flow of capital represented by trade imbalances. Capital does not seem to flow to those nations where it should be most valuable. Instead of capital-rich countries like the United States lending to capital-poor countries, we often observe the opposite. Why is that?

One reason is that there are large differences among nations other than their accumulation of capital. Poor nations have not only lower levels of capital accumulation per worker (represented by *K/L*) but also inferior production capabilities (represented by the variable *A*). For example, compared to rich nations, poor nations may have less access to advanced technologies, lower levels of education (or *human capital*), or less efficient economic policies. Such differences could mean less output for given inputs of capital and labor; in the Cobb–Douglas production function, this is translated into a lower value of the parameter *A*. If so, then capital may not be more valuable in poor nations, even though capital is scarce.

A second reason capital might not flow to poor nations is that property rights are often not enforced. Corruption is much more prevalent; revolutions, coups, and expropriation of wealth are more common; and governments often default on their debts. So even if capital is more valuable in poor nations, foreigners may avoid investing their wealth there simply because they are afraid of losing it. Moreover, local investors face similar incentives. Imagine that you live in a poor nation and are lucky enough to have some wealth to invest; you might well decide that putting it in a safe country like the United States is your best option, even if capital is less valuable there than in your home country.

Whichever of these two reasons is correct, the challenge for poor nations is to find ways to reverse the situation. If these nations offered the same production efficiency and legal protections as the U.S. economy, the direction of international capital flows would likely reverse. The U.S. trade deficit would become a trade surplus, and capital would flow to these emerging nations. Such a change would help the poor of the world escape poverty.[2] ■

6-3 Exchange Rates

Having examined the international flows of capital and of goods and services, we now extend the analysis by considering the prices that apply to these transactions. The *exchange rate* between two countries is the price at which residents of those countries trade with each other. In this section we first examine precisely what the exchange rate measures and then discuss how exchange rates are determined.

Nominal and Real Exchange Rates

Economists distinguish between two exchange rates: the nominal exchange rate and the real exchange rate. Let's discuss each in turn and see how they are related.

The Nominal Exchange Rate

The **nominal exchange rate** is the relative price of the currencies of two countries. For example, if the exchange rate between the U.S. dollar and the Japanese yen is 100 yen per dollar, then you can exchange one dollar for 100 yen in world markets for foreign currency. A Japanese who wants to obtain dollars would pay 100 yen for each dollar he bought. An American who wants to obtain yen would get 100 yen for each dollar he paid. When people refer to "the exchange rate" between two countries, they usually mean the nominal exchange rate.

Notice that an exchange rate can be reported in two ways. If one dollar buys 100 yen, then one yen buys 0.01 dollar. We can say the exchange rate is 100 yen per dollar, or we can say the exchange rate is 0.01 dollar per yen. Because 0.01 equals 1/100, these two ways of expressing the exchange rate are equivalent.

This book always expresses the exchange rate in units of foreign currency per dollar. With this convention, a rise in the exchange rate—say, from 100 to 110 yen per dollar—is called an *appreciation* of the dollar; a fall in the exchange rate is called a *depreciation*. When the domestic currency appreciates, it buys more of the foreign currency; when it depreciates, it buys less. An appreciation is sometimes called a *strengthening* of the currency, and a depreciation is sometimes called a *weakening* of the currency.

The Real Exchange Rate

The **real exchange rate** is the relative price of the goods of two countries. That is, the real exchange rate tells us the rate at which we can trade the goods of one country for the goods of another. The real exchange rate is sometimes called the *terms of trade*.

To see the relation between the real and nominal exchange rates, consider a single good produced in many countries: cars. Suppose an American car costs $30,000 and a similar Japanese car costs 6,000,000 yen. To compare the prices of the two cars, we must convert them into a common currency. If a dollar is worth 100 yen, then the American car costs $100 \times 30{,}000$, or 3,000,000 yen. Comparing the price of the American car (3,000,000 yen) and the price of the Japanese car (6,000,000 yen), we conclude that the American car costs one-half of what the Japanese car costs. In other words, at current prices, we can exchange two American cars for one Japanese car.

We can summarize our calculation as follows:

$$\begin{aligned}\text{Real Exchange Rate} &= \frac{(100 \text{ Yen / Dollar}) \times (30{,}000 \text{ Dollars / American Car})}{(6{,}000{,}000 \text{ Yen / Japanese Car})} \\ &= 0.5\ \frac{\text{Japanese Car}}{\text{American Car}}.\end{aligned}$$

At these prices and this exchange rate, we obtain one-half of a Japanese car per American car. More generally, we can write this calculation as

$$\text{Real Exchange Rate} = \frac{\text{Nominal Exchange Rate} \times \text{Price of Domestic Good}}{\text{Price of Foreign Good}}.$$

The rate at which we exchange foreign and domestic goods depends on the prices of the goods in the local currencies and on the rate at which the currencies are exchanged.

This calculation of the real exchange rate for a single good suggests how we should define the real exchange rate for a broader basket of goods. Let *e* be the nominal exchange rate (the number of yen per dollar), *P* be the price level in the United States (measured in dollars), and P^* be the price level in Japan (measured in yen). Then the real exchange rate ϵ is

$$\begin{aligned}\text{Real Exchange Rate} &= \text{Nominal Exchange Rate} \times \text{Ratio of Price Levels} \\ \epsilon &= e \times (P/P^*).\end{aligned}$$

$$\begin{array}{cccccc} \text{Real Exchange Rate} & = & \text{Nominal Exchange Rate} & \times & \text{Ratio of Price Levels} \\ \epsilon & = & e & \times & (P/P^*)\,. \end{array}$$

The real exchange rate between two countries is computed from the nominal exchange rate and the price levels in the two countries. *If the real exchange rate is high, foreign goods are relatively cheap, and domestic goods are relatively expensive. If the real exchange rate is low, foreign goods are relatively expensive, and domestic goods are relatively cheap*.

The Real Exchange Rate and the Trade Balance

Rip Matteson/The New Yorker/Conde Nast/The Cartoon Bank

"How about Nebraska? The dollar's still strong in Nebraska."

What macroeconomic influence does the real exchange rate exert? To answer this question, remember that the real exchange rate is nothing more than a relative price. Just as the relative price of hamburgers and pizza determines which you choose for lunch, the relative price of domestic and foreign goods affects the demand for these goods.

For example, suppose that the real exchange rate for the United States is low. In this case, because American goods are relatively cheap, Americans will purchase fewer imported goods: they will buy Fords rather than Toyotas, drink Budweiser rather than Heineken, and vacation in Florida rather than in Italy. For the same reason, foreigners will purchase many American goods. They will buy Fords, drink Bud, fly overseas to holiday in Orlando. Because of the actions of both Americans and foreigners, U.S. net exports will be high.

The opposite occurs if the real exchange rate for the United States is high. In this case, American goods are expensive relative to foreign goods. Americans will buy many imported goods, and foreigners will buy few American goods. Therefore, U.S. net exports will be low.

We write this relationship between the real exchange rate and net exports as

$$NX = NX(\epsilon).$$

This equation states that net exports are a function of the real exchange rate. Figure 6-7 illustrates the negative relationship between the trade balance and the real exchange rate.

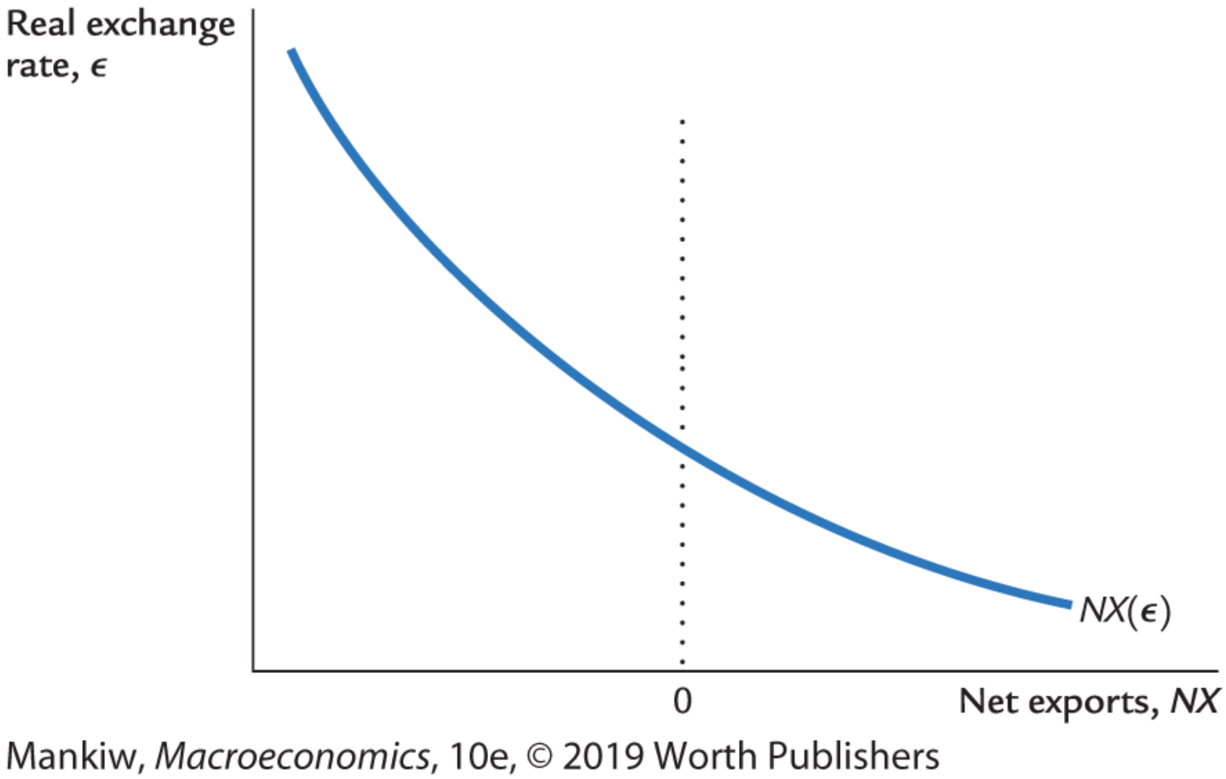


FIGURE 6-7 Net Exports and the Real Exchange Rate The figure shows the relationship between the real exchange rate and net exports: the lower the real exchange rate, the less expensive are domestic goods relative to foreign goods, and thus the greater are a country's net exports. Note that a portion of the horizontal axis measures negative values of *NX*: because imports can exceed exports, net exports can be less than zero.

The Determinants of the Real Exchange Rate

We now have all the pieces needed to construct a model that explains what factors determine the real exchange rate. In particular, we combine the relationship between net exports and the real exchange rate just discussed with the model of the trade balance developed earlier in the chapter. We can summarize the analysis as follows:

- The real value of a currency is inversely related to net exports. When the real exchange rate is lower, domestic goods are less expensive relative to foreign goods, and net exports are greater.

- The trade balance (net exports) must equal the net capital outflow, which in turn equals saving minus investment. Saving is fixed by the consumption function and fiscal policy; investment is fixed by the investment function and the world interest rate.

Figure 6-8 depicts these two conditions. The line showing the relationship between net exports and the real exchange rate slopes downward because a low real exchange rate makes domestic goods relatively inexpensive. The line representing the excess of saving over investment, S–I, $S - I$, is vertical because neither saving nor investment depends on the real exchange rate. The crossing of these two lines determines the equilibrium real exchange rate.

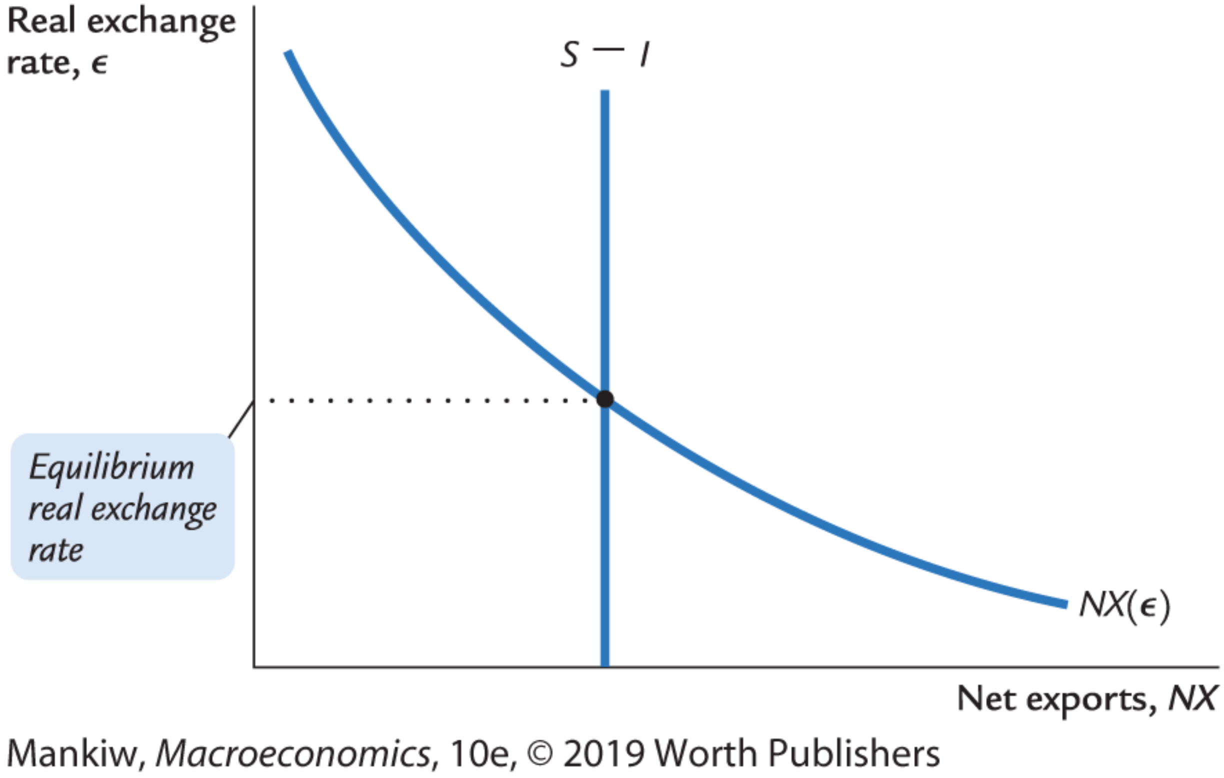


FIGURE 6-8 How the Real Exchange Rate Is Determined The real exchange rate is determined by the intersection of the vertical line representing saving minus investment and the downward-sloping net-exports schedule. At this intersection, the quantity of dollars supplied for the flow of capital abroad equals the quantity of dollars demanded for the net export of goods and services.

Figure 6-8 looks like an ordinary supply-and-demand diagram. In fact, you can think of this diagram as representing the supply and demand for foreign-currency exchange. The vertical line, S–I, $S - I$, represents the net capital outflow and thus the supply of dollars to be exchanged into foreign currency and invested abroad. The downward-sloping line, NX(ϵ), $NX(\epsilon)$, represents the net demand for dollars coming from foreigners who want dollars to buy goods from this country. *At the equilibrium real exchange rate, the supply of dollars available from the net capital outflow balances the demand for dollars by foreigners buying this country’s net exports.*

How Policies Influence the Real Exchange Rate

We can use this model to show how the changes in economic policy we discussed earlier affect the real exchange rate.

Fiscal Policy at Home

What happens to the real exchange rate if the government reduces national saving by increasing government purchases or cutting taxes? As we discussed earlier, this reduction in saving lowers $S - I$ and thus *NX*. That is, the reduction in saving causes a trade deficit.

[Figure 6-9](#) shows how the equilibrium real exchange rate adjusts to ensure that *NX* falls. The change in policy shifts the vertical $S - I$ line to the left, lowering the supply of dollars to be invested abroad. The lower supply causes the equilibrium real exchange rate to rise from ϵ_1 to ϵ_2—that is, the dollar becomes more valuable. Because of the rise in the value of the dollar, domestic goods become more expensive relative to foreign goods, causing exports to fall and imports to rise. These changes in exports and imports both act to reduce net exports.

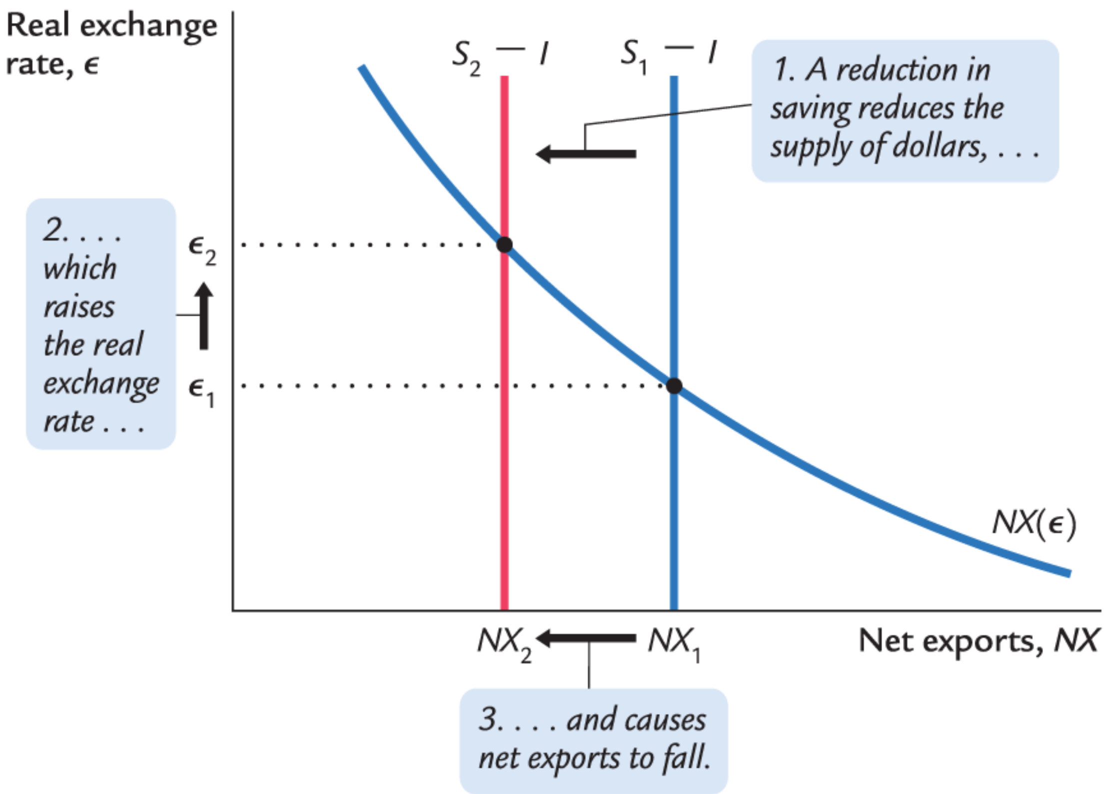


FIGURE 6-9 The Impact of Expansionary Fiscal Policy at Home on the Real Exchange Rate Expansionary fiscal policy at home, such as an increase in government purchases or a cut in taxes, reduces national saving. The fall in saving reduces the supply of dollars to be exchanged into foreign currency, from $S_1 - I$ to $S_2 - I$. This

shift raises the equilibrium real exchange rate from ϵ_1 to ϵ_2.

Fiscal Policy Abroad

What happens to the real exchange rate if foreign governments increase government purchases or cut taxes? Either change in fiscal policy reduces world saving and raises the world interest rate. The increase in the world interest rate reduces domestic investment I, which raises $S - I$ and thus NX. That is, the increase in the world interest rate causes a trade surplus.

Figure 6-10 shows that this change in policy shifts the vertical $S - I$ line to the right, raising the supply of dollars to be invested abroad. The equilibrium real exchange rate falls. That is, the dollar becomes less valuable, and domestic goods become less expensive relative to foreign goods.

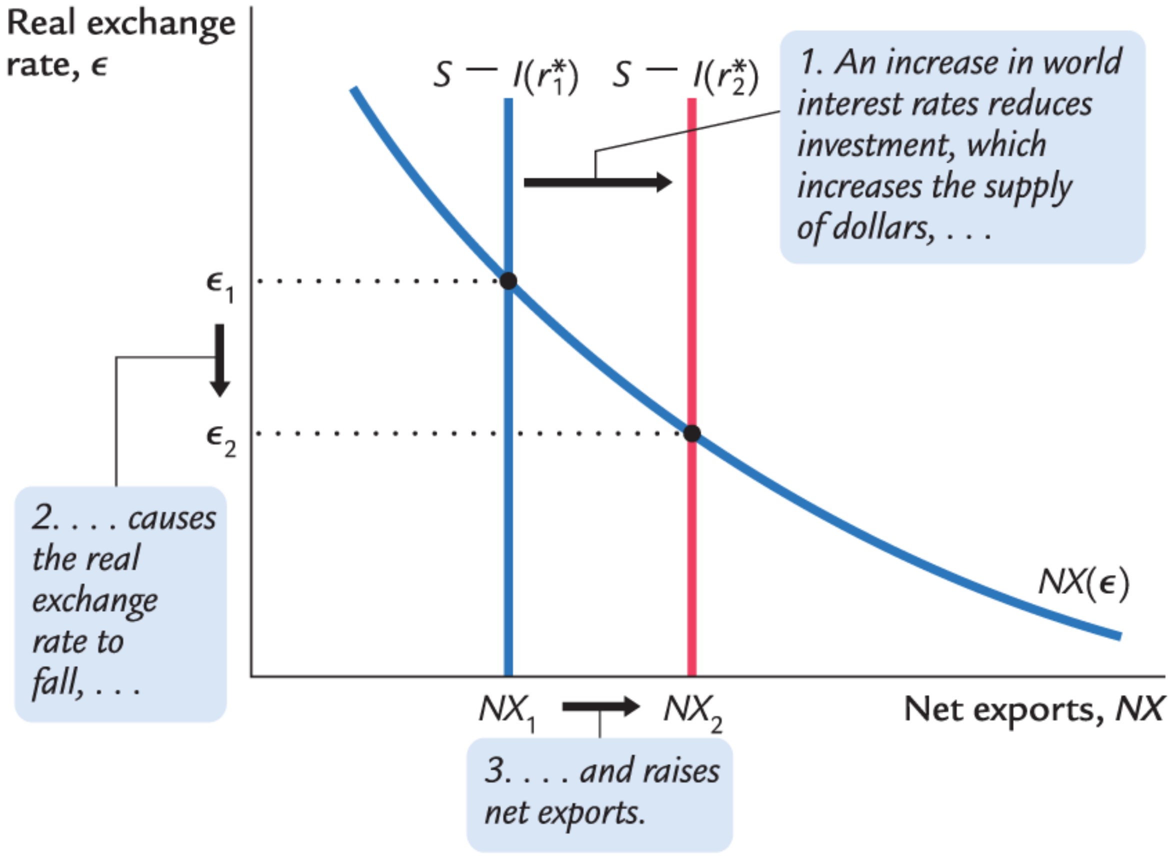
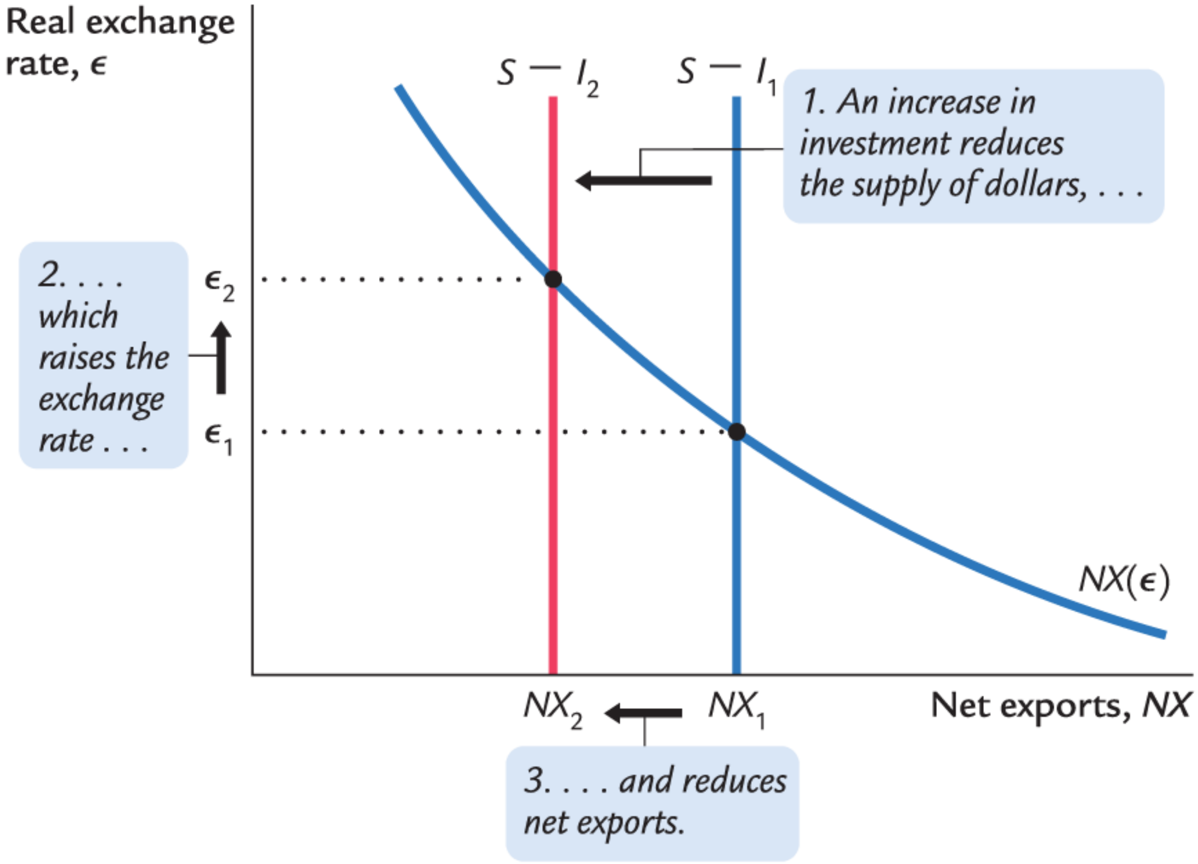


FIGURE 6-10 The Impact of Expansionary Fiscal Policy Abroad on the Real Exchange Rate Expansionary fiscal policy abroad reduces world saving and raises the world interest rate from r_1^* to r_2^*. The increase in the world interest rate reduces investment at home, which in turn raises the supply of dollars to be exchanged into foreign currencies. As a result, the equilibrium real exchange rate falls from ϵ_1 to ϵ_2.

Shifts in Investment Demand

What happens to the real exchange rate if investment demand at home increases, perhaps because Congress passes an investment tax credit? At the given world interest rate, the increase in investment demand leads to higher investment. A higher value of I means lower values of $S - I$ and NX. That is, the increase in investment demand causes a trade deficit.

Figure 6-11 shows that the increase in investment demand shifts the vertical $S - I$ line to the left, reducing the supply of dollars to be invested abroad. The equilibrium real exchange rate rises. Hence, when the investment tax credit makes investing in the United States more attractive, it also increases the value of the U.S. dollars necessary to make these investments. When the dollar appreciates, domestic goods become more expensive relative to foreign goods, and net exports fall.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 6-11 The Impact of an Increase in Investment Demand on the Real Exchange Rate An increase in investment demand raises the quantity of domestic investment from I_1 to I_2. As a result, the supply of dollars to be exchanged into foreign currencies falls from $S - I_1$ to $S - I_2$. This fall in supply raises the equilibrium real exchange rate from ϵ_1 to ϵ_2.

The Effects of Trade Policies

Now that we have a model that explains the trade balance and the real exchange rate, we have the tools to examine the macroeconomic effects of trade policies. Trade policies, broadly defined, are policies designed to directly influence the amount of goods and services exported or imported. Most often, trade policies take the form of protecting domestic industries from foreign competition—either by placing a tax on foreign imports (a tariff) or restricting the amount of goods and services that can be imported (a quota).

For an example of a protectionist trade policy, consider what would happen if the government prohibited the import of foreign cars. For any given real exchange rate, imports would now be lower, implying that net exports (exports minus imports) would be higher. Thus, the net-exports schedule would shift outward, as in Figure 6-12. To see the effects of the policy, we compare the old equilibrium and the new equilibrium. In the new equilibrium, the real exchange rate is higher, and net exports are unchanged. Despite the shift in the net-exports schedule, the equilibrium level of net exports remains the same, because the protectionist policy does not alter either saving or investment.

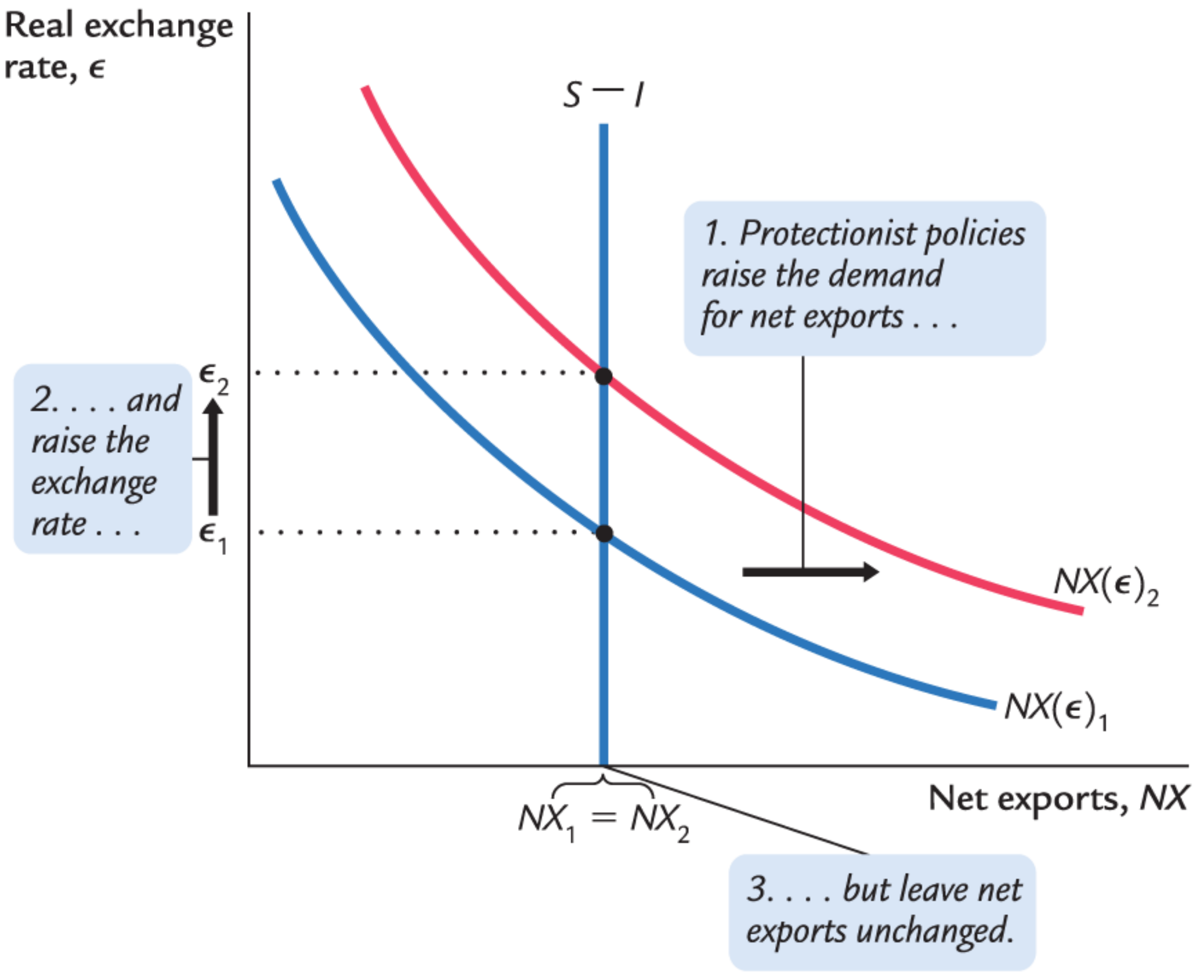


FIGURE 6-12 The Impact of Protectionist Trade Policies on the Real Exchange Rate A protectionist trade policy, such as a ban on imported cars, shifts the net-exports schedule from NX(ϵ)1$NX(\epsilon)_1$ to NX(ϵ)2,$NX(\epsilon)_2$, which raises the real exchange rate from ϵ1ϵ_1 to ϵ2.ϵ_2. Notice that, despite the shift in the net-exports schedule, the equilibrium level of net exports is unchanged.

This analysis shows that protectionist trade policies do not affect the trade balance. This surprising conclusion is often overlooked in the popular debate over trade policies. Because a trade deficit reflects an excess of imports over exports, one might guess that reducing imports—such as by prohibiting the import of foreign cars—would reduce a trade deficit. Yet our model shows that protectionist policies lead only to an appreciation of the real exchange rate. The increase in the price of domestic goods relative to foreign goods tends to lower net exports by stimulating imports and depressing exports. Thus, the appreciation offsets the increase in net exports that is directly attributable to the trade restriction.

Although protectionist trade policies do not alter the trade balance, they do affect the amount of trade. As we have seen, because the real exchange rate appreciates, the goods and services a country produces become more expensive relative to foreign goods and services. The country therefore exports less in the new equilibrium. Because net exports are unchanged, it must import less as well. (The appreciation of the exchange rate does stimulate imports to some extent, but this only partly offsets the decrease in imports due to the trade restriction.) Thus, protectionist policies reduce both imports and exports.

This fall in the amount of trade is the reason economists usually oppose protectionist policies. International trade benefits all countries by allowing each country to specialize in what it produces best and by providing each country with a greater variety of goods and services. Protectionist policies diminish these gains from trade. Although these policies benefit certain groups within society—for example, a ban on imported cars helps domestic car producers—society on average is worse off when policies reduce the amount of international trade.

The Determinants of the Nominal Exchange Rate

Having seen what determines the real exchange rate, we now turn our attention to the nominal exchange rate—the rate at which the currencies of two countries trade. Recall the relationship between the real and the nominal exchange rate:

$$\begin{array}{cccccc} \text{Real Exchange Rate} & = & \text{Nominal Exchange Rate} & \times & \text{Ratio of Price Levels} \\ \epsilon & = & e & \times & (P/P^*). \end{array}$$

We can write the nominal exchange rate as

$$e = \epsilon \times (P^*/P).$$

This equation shows that the nominal exchange rate depends on the real exchange rate and the price levels in the two countries. Given the value of the real exchange rate, if the domestic price level P rises, then the nominal exchange rate e will fall: because a dollar is worth less, a dollar will buy fewer yen. However, if the Japanese price level P^* rises, then the nominal exchange rate will increase: because the yen is worth less, a dollar will buy more yen.

It is instructive to consider changes in exchange rates over time. The exchange rate equation can be written

$$\%\text{ Change in } e = \%\text{ Change in } \epsilon + \%\text{ Change in } P^* - \%\text{ Change in } P.$$

The percentage change in ϵ is the change in the real exchange rate. The percentage change in *P* is the domestic inflation rate π, and the percentage change in P^* is the foreign country's inflation rate π^* Thus, the percentage change in the nominal exchange rate is

$$\begin{array}{ccccc} \%\text{ Change in } e & = & \%\text{ Change in } \epsilon & + & (\pi^* - \pi) \\ \text{Percentage Change in Nominal Exchange Rate} & = & \text{Percentage Change in Real Exchange Rate} & + & \text{Difference in Inflation Rates} \end{array}$$

This equation states that the percentage change in the nominal exchange rate between the currencies of two countries equals the percentage change in the real exchange rate plus the difference in their inflation rates. *If a country has a high rate of inflation relative to the United States, a dollar will buy an increasing amount of the foreign currency over time. If a country has a low rate of inflation relative to the United States, a dollar will buy a decreasing amount of the foreign currency over time.*

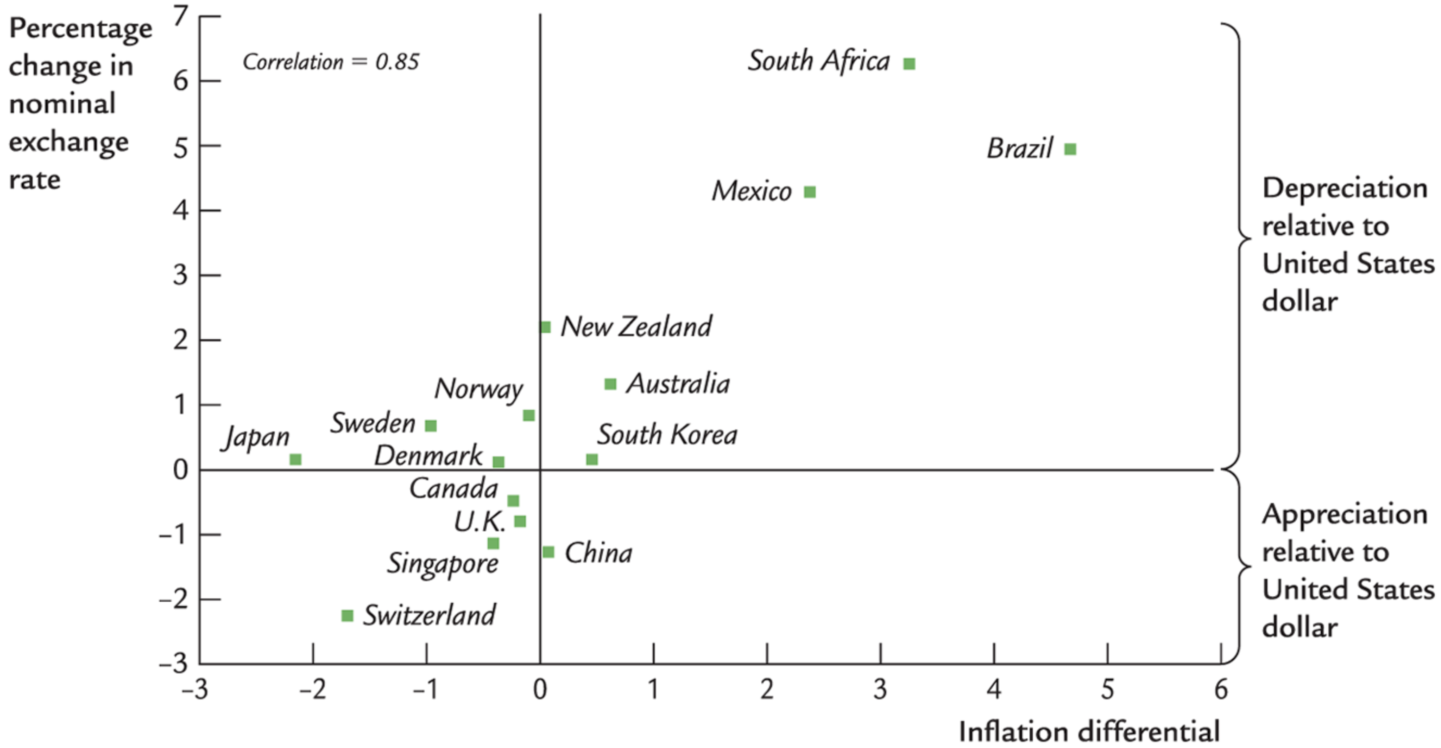
This analysis shows how monetary policy affects the nominal exchange rate. We know from Chapter 5 that high growth in the money supply leads to high inflation. Here, we have just seen that one consequence of high inflation is a depreciating currency: high π implies falling *e*. In other words, just as growth in the amount of money raises the price of goods measured in terms of money, it also tends to raise the price of foreign currencies measured in terms of the domestic currency.

CASE STUDY

Inflation and Nominal Exchange Rates

If we look at data on exchange rates and price levels of different countries, we quickly see the importance of inflation for explaining changes in the nominal exchange rate. The most dramatic examples come from periods of very high inflation. For example, the price level in Mexico rose by 2,300 percent from 1983 to 1988. Because of this inflation, the number of pesos a person could buy with a U.S. dollar rose from 144 in 1983 to 2,281 in 1988.

The same relationship holds true for countries with more moderate inflation. Figure 6-13 is a scatterplot showing the relationship between inflation and the exchange rate for 15 countries. On the horizontal axis is the difference between each country's average inflation rate and the average inflation rate of the United States $(\pi^* - \pi)$. On the vertical axis is the average percentage change in the exchange rate between each country’s currency and the U.S. dollar (percentage change in *e*). The positive relationship between these two variables is clear in this figure. The correlation between these variables—a statistic that runs from -1 to $+1$ and measures how closely the variables are related—is 0.85. Countries with relatively high inflation tend to have depreciating currencies (you can buy more of them with your dollars over time), and countries with relatively low inflation tend to have appreciating currencies (you can buy less of them with your dollars over time).



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 6-13 Inflation Differentials and the Exchange Rate This scatterplot shows the relationship between inflation and the nominal exchange rate. The horizontal axis shows the country’s average inflation rate minus the U.S. average inflation rate over the period 2000–2016. The vertical axis is the average percentage change in the country’s exchange rate (per U.S. dollar) over that period. This figure shows that countries with relatively high inflation tend to have depreciating currencies and that countries with relatively low inflation tend to have appreciating currencies.

Data from: St. Louis FRED.

For example, consider the exchange rate between Swiss francs and U.S. dollars. Both Switzerland and the United States have experienced inflation over these years, so both the franc and the dollar buy fewer goods than they once did. But, as Figure 6-13 shows, inflation in Switzerland has been lower than inflation in the United States. This means that the value of the franc has fallen less than the value of the dollar. Therefore, the number of Swiss francs you can buy with a U.S. dollar has been falling over time. ■

The Special Case of Purchasing-Power Parity

A famous hypothesis in economics, called the *law of one price*, states that the same good cannot sell for different prices in different locations at the same time. If a bushel of wheat sold for less in New York than in Chicago, it would be profitable to buy wheat in New York and then sell it in Chicago. This profit opportunity would become quickly apparent to astute arbitrageurs—people who specialize in “buying low” in one market and “selling high” in another. As the arbitrageurs took advantage of this opportunity, they would increase the demand for wheat in New York and increase the supply of wheat in Chicago. Their actions would drive the price up in New York and down in Chicago, thereby ensuring that prices are equalized in the two markets.

The law of one price applied to the international marketplace is called **purchasing-power parity**. It states that if international arbitrage is possible, then a dollar (or any other currency) must have the same purchasing power in every country. The argument goes as follows. If a dollar could buy more wheat domestically than abroad, there would be opportunities to profit by buying wheat domestically and selling it abroad. Profit-seeking arbitrageurs would drive up the domestic price of wheat relative to the foreign price. Similarly, if a dollar could buy more wheat abroad than domestically, the arbitrageurs would buy wheat abroad and sell it domestically, driving down the domestic price relative to the foreign price. Thus, profit-seeking by international arbitrageurs causes wheat prices to be the same in all countries.

We can interpret purchasing-power parity using our model of the real exchange rate. The quick action of these international arbitrageurs implies that net exports are highly sensitive to small movements in the real exchange rate. A small decrease in the price of domestic goods relative to foreign goods—that is, a small decrease in the real exchange rate—causes arbitrageurs to buy goods domestically and sell them abroad. Similarly, a small increase in the relative price of domestic goods causes arbitrageurs to import goods from abroad. Therefore, as in Figure 6-14, the net-exports schedule is very flat at the real exchange rate that equalizes purchasing power among countries: any small movement in the real exchange rate leads to a large change in net exports. This extreme sensitivity of net exports guarantees that the equilibrium real exchange rate is always close to the level that ensures purchasing-power parity.

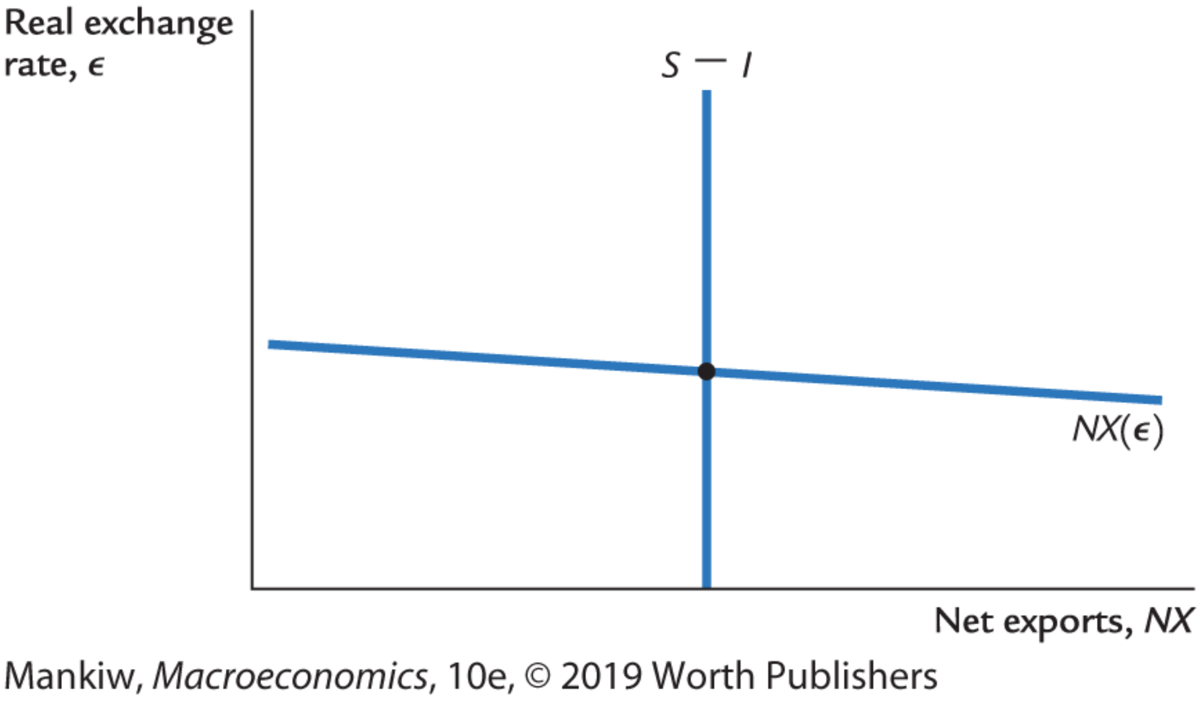


FIGURE 6-14 Purchasing-Power Parity The law of one price applied to the international marketplace suggests that net exports are highly sensitive to small movements in the real exchange rate. This high sensitivity is reflected here with a very flat net-exports schedule.

Purchasing-power parity has two key implications. First, because the net-exports schedule is flat, changes in saving or investment do not affect the real or nominal exchange rate. Second, because the real exchange rate is fixed, all changes in the nominal exchange rate result from changes in price levels.

Is the theory of purchasing-power parity realistic? Most economists believe that, despite its appealing logic, purchasing-power parity is not a completely accurate description of the world. One reason is that many goods and services are not easily traded. A haircut can be more expensive in Tokyo than in New York, but there is no room for international arbitrage because it is impossible to transport haircuts. In addition, even tradable goods are not always perfect substitutes. Because some consumers prefer Toyotas and others prefer Fords, the relative price of Toyotas and Fords can vary to some extent without leaving any profit opportunities. As a result, real exchange rates do in fact vary over time.

Although the theory of purchasing-power parity does not describe the world perfectly, it provides a reason why movement in the real exchange rate will be limited. Its underlying logic is compelling: the farther the real exchange rate drifts from the level predicted by purchasing-power parity, the greater the incentive for people to engage in international arbitrage in goods. We cannot rely on purchasing-power parity to eliminate all changes in the real exchange rate, but this theory leads us to expect that fluctuations in the real exchange rate will typically be small or temporary.[3]

CASE STUDY

The Big Mac Around the World

The theory of purchasing-power parity says that after we adjust for exchange rates, goods should sell for the same price everywhere. Conversely, it says that the exchange rate between two currencies should depend on the price levels in the two countries.

To see how well this theory works, *The Economist*, an international news magazine, regularly collects data on the price of a good sold in many countries: the McDonald’s Big Mac hamburger. According to purchasing-power parity, the price of a Big Mac should be closely related to the country’s nominal exchange rate. The higher the price of a Big Mac in the local currency, the higher the exchange rate (measured in units of local currency per U.S. dollar) should be.

Table 6-2 presents the international prices in 2017, when a Big Mac sold for $5.30 in the United States (this was the average price in New York, San Francisco, Chicago, and Atlanta). With these data we can use the theory of purchasing-power parity to predict nominal exchange rates. For example, because a Big Mac cost 16.50 reals in Brazil we would predict that the exchange rate between the dollar and the real was 16.50/5.30, or 3.11, reals per dollar. At this exchange rate, a Big Mac would have cost the same in Brazil and the United States.

TABLE 6-2

Big Mac Prices and the Exchange Rate: An Application of Purchasing-Power Parity

Country	Currency	Price of a Big Mac	Exchange Rate (per U.S. dollar)	
			Predicted	Actual
Indonesia	Rupiah	32126.00	6062	13370

Colombia	Peso	9900.00	1868	3052
South Korea	Won	4400.00	830	1145
Chile	Peso	2550.00	481	663
Hungary	Forint	862.00	163	269
Japan	Yen	380.00	71.7	113.1
Pakistan	Rupee	375.00	70.8	105.2
India	Rupee	178.00	33.6	64.6
Russia	Rouble	137.00	25.8	60.1
Philippines	Peso	134.00	25.3	50.6
Thailand	Baht	119.00	22.5	34.0
Czech Republic	Koruna	75.00	14.2	22.9
Argentina	Peso	70.00	13.2	17.0
Taiwan	NT Dollar	69.00	13.0	30.5
Norway	Kroner	49.00	9.25	8.29
Mexico	Peso	49.00	9.25	17.79
Sweden	Krona	48.97	9.24	8.42
Egypt	Pound	31.37	5.92	17.89
Denmark	D. Krone	30.00	5.66	6.51
South Africa	Rand	30.00	5.66	13.27
China	Yuan	19.80	3.74	6.79
Hong Kong	HK Dollar	19.20	3.62	7.81
Israel	Shekel	16.90	3.19	3.54
Brazil	Real	16.50	3.11	3.23
Saudi Arabia	Riyal	12.00	2.26	3.75
Turkey	Lira	10.75	2.03	3.58
Peru	Sol	10.50	1.98	3.25
Poland	Zloty	10.10	1.91	3.71
Malaysia	Ringgit	8.60	1.62	4.29
Switzerland	S. Franc	6.50	1.23	0.96

New Zealand	NZ Dollar	6.10	1.15	1.38
Canada	C. Dollar	5.97	1.13	1.28
Australia	A. Dollar	5.90	1.11	1.30
Singapore	S. Dollar	5.60	1.06	1.38
United States	Dollar	5.30	1.00	1.00
Euro area	Euro	3.91	0.74	0.88
Britain	Pound	3.19	0.60	0.78

Note: The predicted exchange rate is the exchange rate that would make the price of a Big Mac in that country equal to its price in the United States.

Data from: *The Economist*.

Table 6-2 shows the predicted and actual exchange rates for 36 countries, plus the euro area, ranked by the predicted exchange rate. You can see that the evidence on purchasing-power parity is mixed. As the last two columns show, the actual and predicted exchange rates are usually in the same ballpark. Our theory predicts, for instance, that a U.S. dollar should buy the greatest number of Indonesian rupiahs and fewest British pounds, and this turns out to be true. In the case of Brazil, the predicted exchange rate of 3.11 reals per dollar is close to the actual exchange rate of 3.23. Yet the theory's predictions are far from exact and, in many cases, are off by 30 percent or more. Hence, although the theory of purchasing-power parity provides a rough guide to exchange rates, it does not explain them completely. ■

6-4 Conclusion: The United States as a Large Open Economy

In this chapter we have seen how a small open economy works. We have examined the determinants of the international flow of funds for capital accumulation and the international flow of goods and services. We have also examined the determinants of a country's real and nominal exchange rates. Our analysis shows how various policies—monetary policies, fiscal policies, and trade policies—affect the trade balance and the exchange rate.

The economy we have studied is "small" in the sense that its interest rate is fixed by world financial markets. That is, we have assumed that this economy does not affect the world interest rate and that the economy can borrow and lend at the world interest rate in unlimited amounts. This assumption contrasts with the assumption we made when studying the closed economy in Chapter 3. In the closed economy, the domestic interest rate equilibrates domestic saving and domestic investment, implying that policies that influence saving or investment alter the equilibrium interest rate.

Which of these analyses should we apply to an economy such as that of the United States? The answer is a little of both. The United States is neither so large nor so isolated that it is immune to developments occurring abroad. The large trade deficits of the 1980s, 1990s, and 2000s show the importance of international financial markets for funding U.S. investment. Hence, the closed-economy analysis of Chapter 3 cannot by itself fully explain the impact of policies on the U.S. economy.

Yet the United States is not so small and so open that the analysis of this chapter applies perfectly either. First, the U.S. economy is large enough that it can influence world financial markets. Second, capital may not be perfectly mobile across countries. If individuals prefer holding their wealth in domestic rather than foreign assets, funds for capital accumulation will not flow freely to equate interest rates in all countries. For these two reasons, we cannot directly apply our model of the small open economy to the United States.

When analyzing policy for a country such as the United States, we need to combine the closed-economy logic of Chapter 3 and the small-open-economy logic of this chapter. The appendix to this chapter builds a model of an economy between these two extremes. In this intermediate case, there is international borrowing and lending, but the interest rate is not fixed by world financial markets. Instead, the more the economy borrows from abroad, the higher the interest rate it must offer foreign investors. The results, not surprisingly, are a mixture of the two polar cases we have already examined.

Consider, for example, a reduction in national saving due to a fiscal expansion. As in the closed economy, this policy raises the real interest rate and crowds out domestic investment. As in the small open economy, it also reduces the net capital outflow, leading to a trade deficit and an appreciation of the exchange rate. Hence, although the model of the small open economy examined here does not precisely describe an economy such as that of the United States, it provides approximately the right answer to how policies affect the trade balance and the exchange rate.

APPENDIX

The Large Open Economy

When analyzing policy for a country such as the United States, we need to combine the closed-economy logic of Chapter 3 and the small-open-economy logic of this chapter. This appendix presents a model of an economy between these two extremes, called the *large open economy*.

Net Capital Outflow

The key difference between the small and large open economies is the behavior of the net capital outflow. In the model of the small open economy, capital flows freely into or out of the economy at a fixed world interest rate r^* The model of the large open economy makes a different assumption about international capital flows. To understand this assumption, keep in mind that the net capital outflow is the amount that domestic investors lend abroad minus the amount that foreign investors lend here.

Imagine that you are a domestic investor—such as the portfolio manager of a university endowment—deciding where to invest your funds. You could invest domestically (for example, by making loans to U.S. companies), or you could invest abroad (by making loans to foreign companies). Many factors may affect your decision, but surely one of them is the interest rate you can earn. The higher the interest rate you can earn domestically, the less attractive you would find foreign investment.

Investors abroad face a similar decision. They have a choice between investing in their home country and lending to someone in the United States. The higher the interest rate in the United States, the more willing foreigners are to lend to U.S. companies and to buy U.S. assets.

Thus, because of the behavior of both domestic and foreign investors, the net flow of capital to other countries, which we'll denote as *CF*, is negatively related to the domestic real interest rate *r*. As the interest rate rises, less domestic saving flows abroad, and more funds for capital accumulation flow in from other countries. We write this as

$$CF = CF(r).$$

This equation states that the net capital outflow is a function of the domestic interest rate. Figure 6-15 illustrates this relationship. Notice that *CF* can be either positive or negative, depending on whether the economy is a lender or borrower in world financial markets.

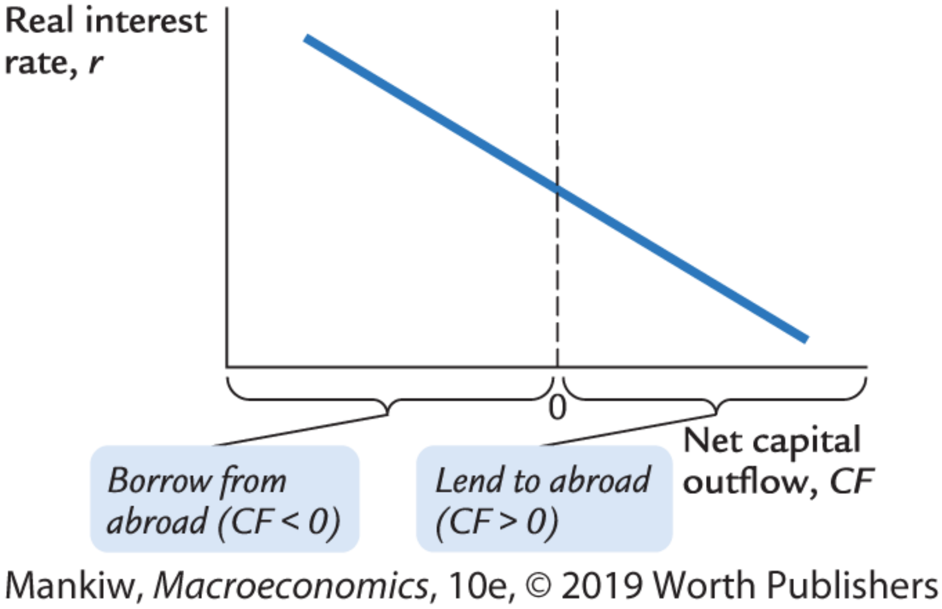


FIGURE 6-15 How the Net Capital Outflow Depends on the Interest Rate A higher domestic interest rate discourages domestic investors from lending abroad and encourages foreign investors to lend here. Therefore, net capital outflow *CF* is negatively related to the interest rate.

To see how this *CF* function relates to our previous models, consider Figure 6-16. This figure shows two special cases: a vertical *CF* function and a horizontal *CF* function.

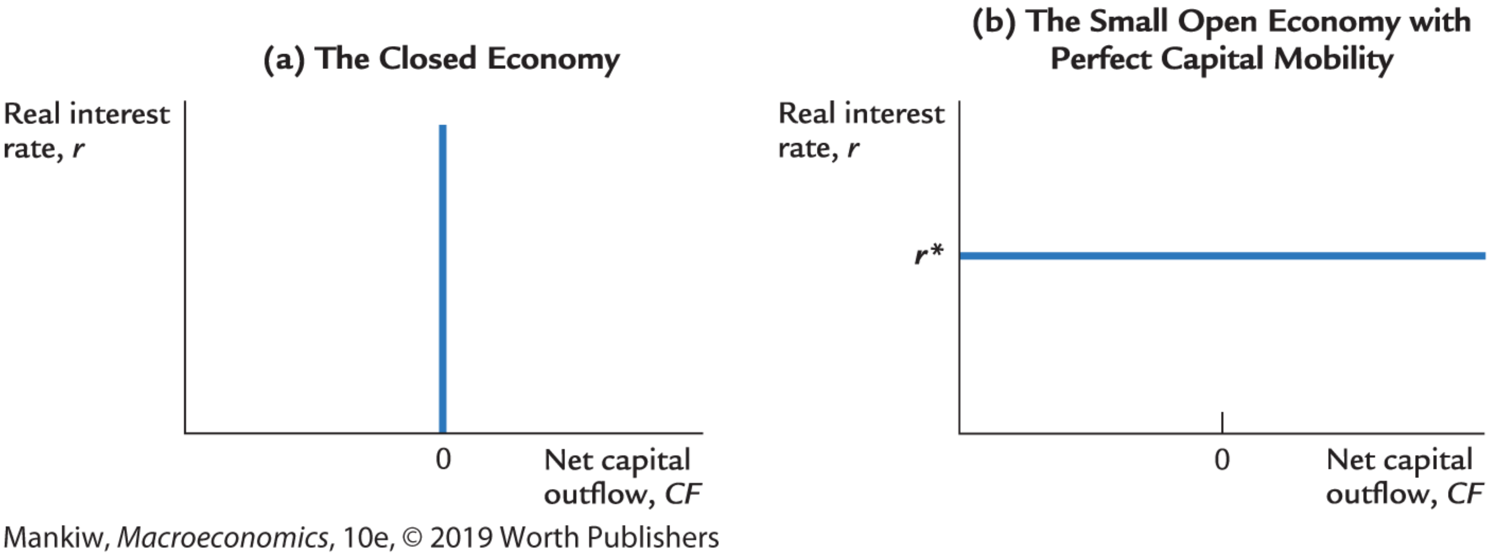


FIGURE 6-16 Two Special Cases In the closed economy, shown in panel (a), the net capital outflow is zero for all interest rates. In the small open economy with perfect capital mobility, shown in panel (b), the net capital outflow is perfectly elastic at the world interest rate r^*.

The closed economy is the special case shown in panel (a) of Figure 6-16. In the closed economy, there is no international borrowing or lending, and the interest rate adjusts to equilibrate domestic saving and investment. This means that $CF = 0$ at all interest rates. This situation would arise if investors here and abroad were unwilling to hold foreign assets, regardless of the return. It might also arise if the government prohibited its citizens from transacting in foreign financial markets, as some governments do.

The small open economy with perfect capital mobility is the special case shown in panel (b) of Figure 6-16. In this case, capital flows freely into and out of the country at the fixed world interest rate r*.r^*. This situation would arise if investors here and abroad bought whatever asset yielded the highest return and if this economy were too small to affect the world interest rate. The economy's interest rate would be fixed at the interest rate prevailing in world financial markets.

Why isn't the interest rate of a large open economy such as the United States fixed by the world interest rate? There are two reasons. The first is that the United States is large enough to influence world financial markets. The more the United States lends abroad, the greater is the supply of loans in the world economy, and the lower interest rates become around the world. The more the United States borrows from abroad (that is, the more negative *CF* becomes), the higher are world interest rates. We use the label "large open economy" because this model applies to an economy large enough to affect world interest rates.

There is, however, a second reason the interest rate in an economy may not be fixed by the world interest rate: capital may not be perfectly mobile. That is, investors here and abroad may prefer to hold their wealth in domestic rather than foreign assets. Such a preference for domestic assets could arise because of imperfect information about foreign assets or because of government impediments to international borrowing and lending. In either case, funds for capital accumulation will not flow freely to equalize interest rates in all countries. Instead, the net capital outflow will depend on domestic interest rates relative to foreign interest rates. U.S. investors will lend abroad only if U.S. interest rates are comparatively low, and foreign investors will lend in the United States only if U.S. interest rates are comparatively high. The large-open-economy model, therefore, may apply even to a small economy if capital does not flow freely into and out of the economy.

Hence, either because the large open economy affects world interest rates, or because capital is imperfectly mobile, or perhaps for both reasons, the *CF* function slopes downward. Except for this new downward-sloping *CF* function, the model of the large open economy resembles the model of the small open economy. We put all the pieces together in the next section.

The Model

To understand how the large open economy works, we need to consider two key markets: the market for loanable funds (where the interest rate is determined) and the market for foreign exchange (where the exchange rate is determined). The interest rate and the exchange rate are two prices that guide the allocation of resources.

The Market for Loanable Funds

An open economy's saving *S* is used in two ways: to finance domestic investment *I* and to finance the net capital outflow *CF*. We can write

$$S = I + CF.$$

Consider how these three variables are determined. National saving is fixed by the level of output, fiscal policy, and the consumption function. Investment and net capital outflow both depend on the domestic real interest rate. We can write

$$\overline{S} = I(r) + CF(r).$$

Figure 6-17 shows the market for loanable funds. The supply of loanable funds is national saving. The demand for loanable funds is the sum of the demand for domestic investment and the demand for foreign investment (net capital outflow). The interest rate adjusts to equilibrate supply and demand.

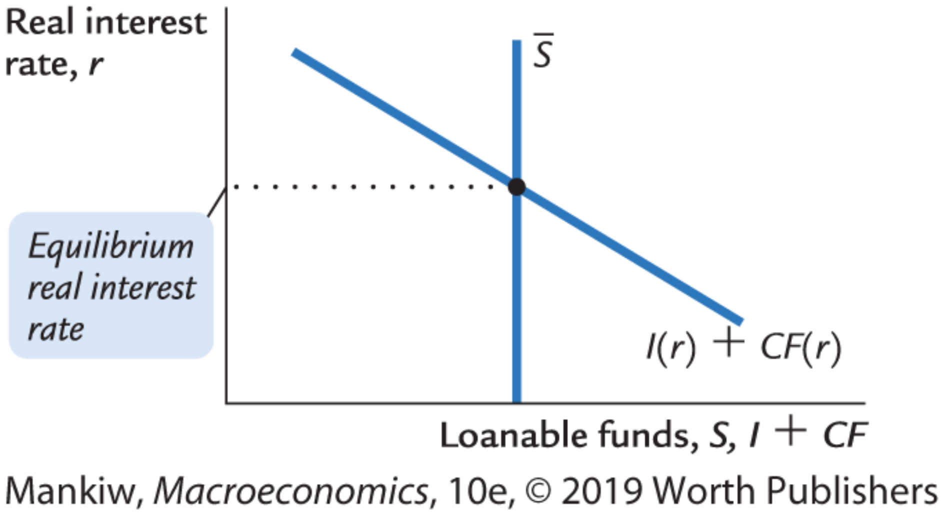


FIGURE 6-17 The Market for Loanable Funds in the Large Open Economy At the equilibrium interest rate, the supply of loanable funds from saving *S* balances the demand for loanable funds from domestic investment *I* and capital investments abroad *CF*.

The Market for Foreign Exchange

Next, consider the relationship between the net capital outflow and the trade balance. The national income accounts identity tells us

$$NX = S - I.$$

Because *NX* is a function of the real exchange rate, and because $CF = S - I,$ we can write

$$NX(\epsilon) = CF.$$

Figure 6-18 shows the equilibrium in the market for foreign exchange. Once again, the real exchange rate is the price that equilibrates the trade balance and the net capital outflow.

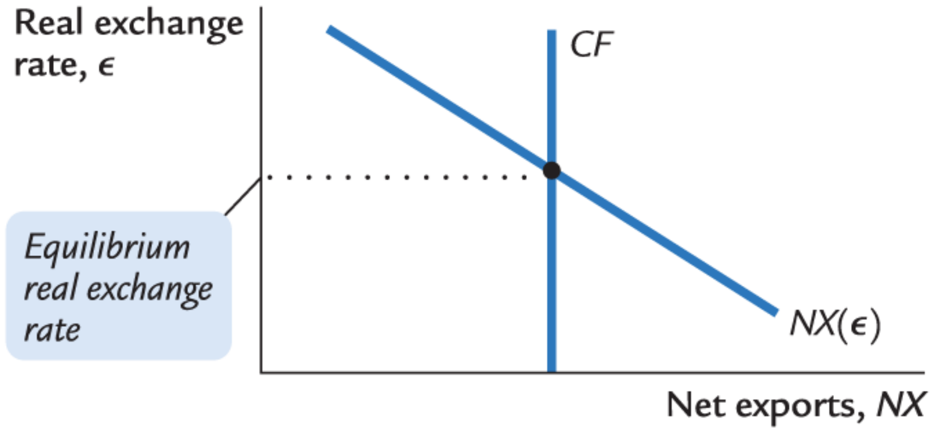


FIGURE 6-18 The Market for Foreign-Currency Exchange in the Large Open Economy At the equilibrium exchange rate, the supply of dollars from the net capital outflow, *CF*, balances the demand for dollars from a country's net exports of goods and services, *NX*.

The last variable we should consider is the nominal exchange rate. As before, the nominal exchange rate is the real exchange rate times the ratio of the price levels:

$$e = \epsilon \times (P^*/P).$$

The real exchange rate is determined as in Figure 6-18, and the price levels are determined by monetary policies here and abroad, as we discussed in Chapter 5. Forces that move the real exchange rate or the price levels also move the nominal exchange rate.

Policies in the Large Open Economy

We can now consider how economic policies influence the large open economy. Figure 6-19 shows the three diagrams we need for the analysis. Panel (a) shows the equilibrium in the market for loanable funds; panel (b) shows the relationship between the equilibrium interest rate and the net capital outflow; and panel (c) shows the equilibrium in the market for foreign exchange.

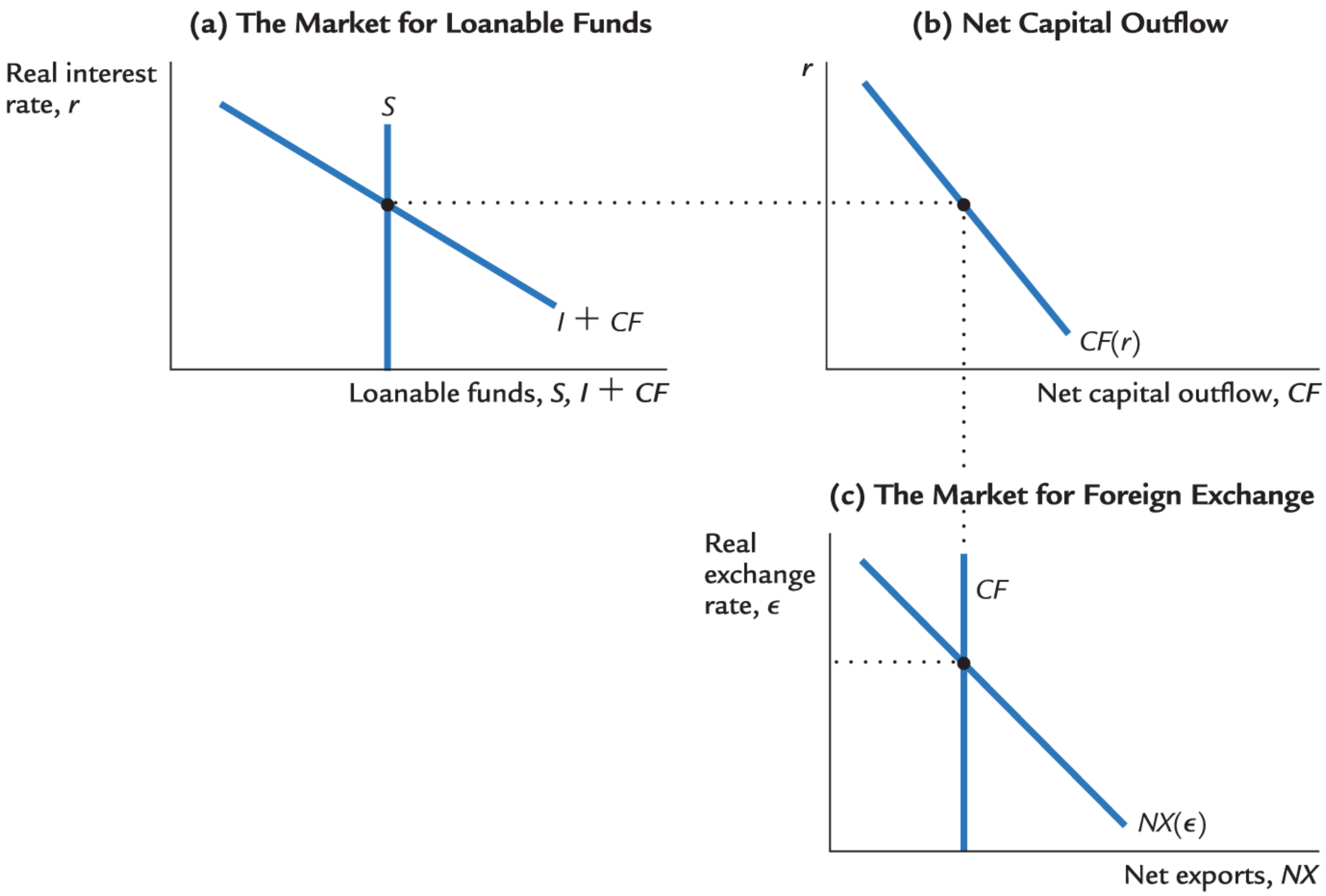


FIGURE 6-19 The Equilibrium in the Large Open Economy Panel (a) shows that the market for loanable funds determines the equilibrium interest rate. Panel (b) shows that the interest rate determines the net capital outflow, which in turn determines the supply of dollars to be exchanged into foreign currency. Panel (c) shows that the real exchange rate adjusts to balance this supply of dollars with the demand coming from net exports.

Fiscal Policy at Home

Consider the effects of expansionary fiscal policy—an increase in government purchases or a decrease in taxes. Figure 6-20 shows what happens. The policy reduces national saving S, thereby reducing the supply of loanable funds and raising the equilibrium interest rate r. The higher interest rate reduces both domestic investment I and the net capital outflow CF. The fall in the net capital outflow reduces the supply of dollars to be exchanged into foreign currency. The exchange rate appreciates, and net exports fall.

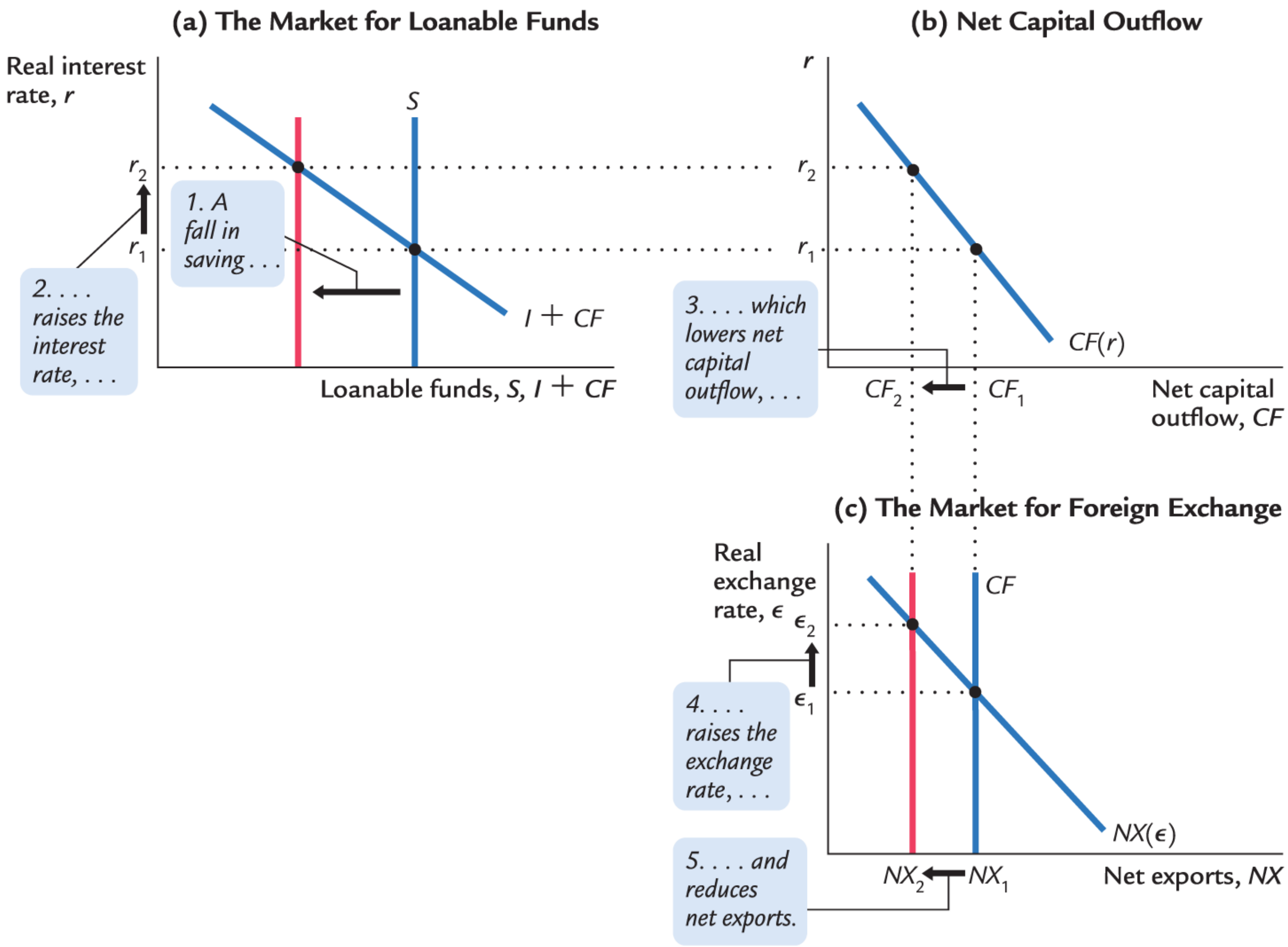


FIGURE 6-20 A Reduction in National Saving in the Large Open Economy Panel (a) shows that a reduction in national saving lowers the supply of loanable funds. The equilibrium interest rate rises. Panel (b) shows that the higher interest rate lowers the net capital outflow. Panel (c) shows that the reduced capital outflow means a reduced supply of dollars in the market for foreign-currency exchange. The reduced supply of dollars causes the real exchange rate to appreciate and net exports to fall.

Note that the impact of fiscal policy in this model combines its impact in the closed economy and its impact in the small open economy. As in the closed economy, a fiscal expansion in a large open economy raises the interest rate and crowds out investment. As in the small open economy, a fiscal expansion causes a trade deficit and an appreciation in the exchange rate.

One way to see how the three types of economy are related is to consider the identity

$$S = I + NX.$$

In all three cases, expansionary fiscal policy reduces national saving S. In the closed economy, the fall in S coincides with an equal fall in I, and NX stays constant at zero. In the small open economy, the fall in S

coincides with an equal fall in *NX*, and *I* remains constant at the level fixed by the world interest rate. The large open economy is the intermediate case: both *I* and *NX* fall, each by less than the fall in *S*.

Shifts in Investment Demand

Suppose that the investment demand schedule shifts outward, perhaps because Congress passes an investment tax credit. Figure 6-21 shows the effect. The demand for loanable funds rises, raising the equilibrium interest rate. The higher interest rate reduces the net capital outflow: Americans make fewer loans abroad, and foreigners make more loans to Americans. The fall in the net capital outflow reduces the supply of dollars in the market for foreign exchange. The exchange rate appreciates, and net exports fall.

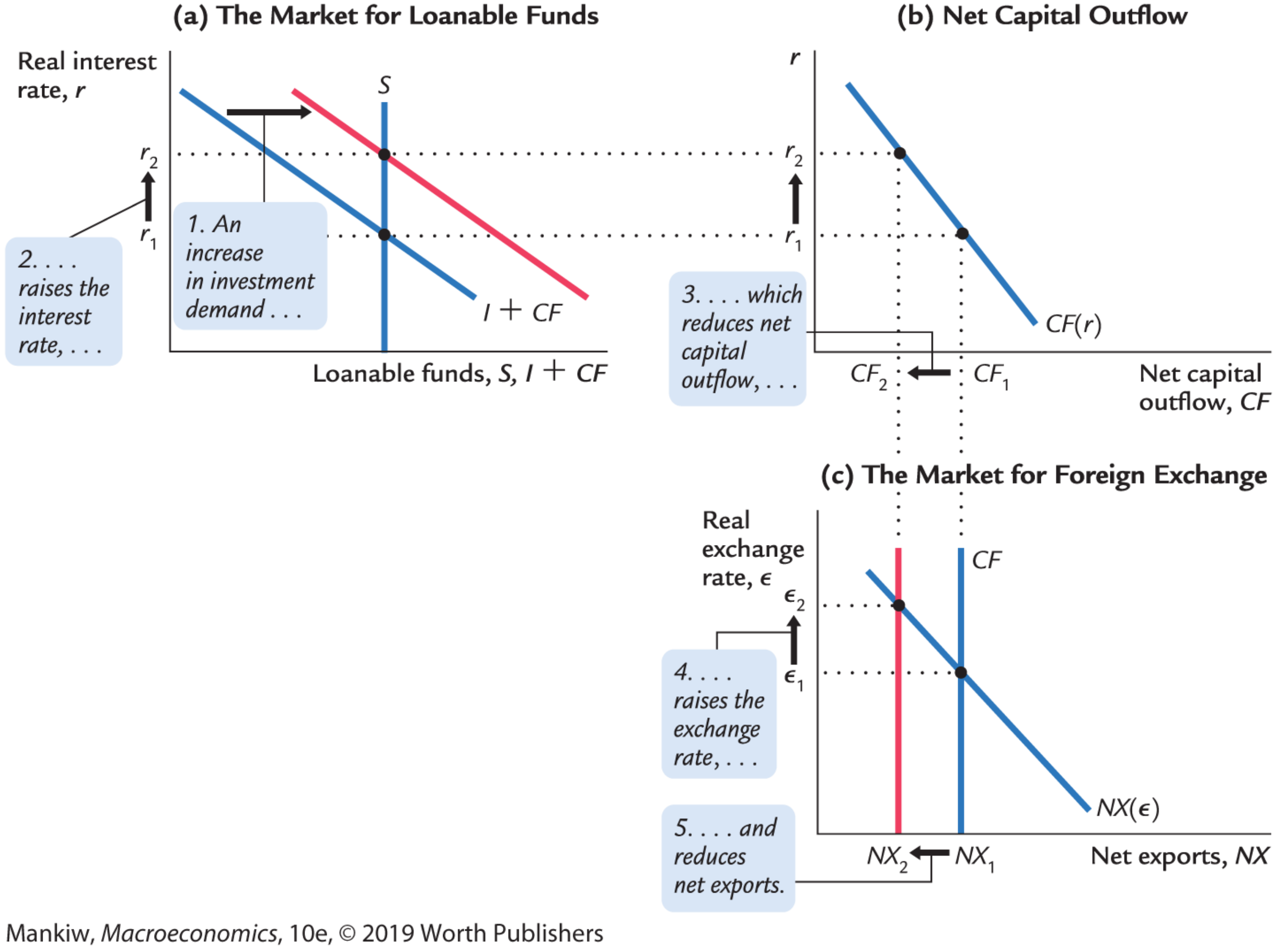


FIGURE 6-21 An Increase in Investment Demand in the Large Open Economy Panel (a) shows that an increase in investment demand raises the interest rate. Panel (b) shows that the higher interest rate lowers the net capital outflow. Panel (c) shows that a lower capital outflow causes the real exchange rate to appreciate and net exports to fall.

Trade Policies

Figure 6-22 shows the effect of a trade restriction, such as an import quota. The reduced demand for imports shifts the net exports schedule outward in panel (c). Because nothing has changed in the market for loanable funds, the interest rate remains the same, implying that the net capital outflow remains the same. The shift in the net-exports schedule causes the dollar to appreciate in the market for foreign exchange. The appreciation makes U.S. goods more expensive relative to foreign goods, which depresses exports and stimulates imports. In the end, the trade restriction does not affect the trade balance.

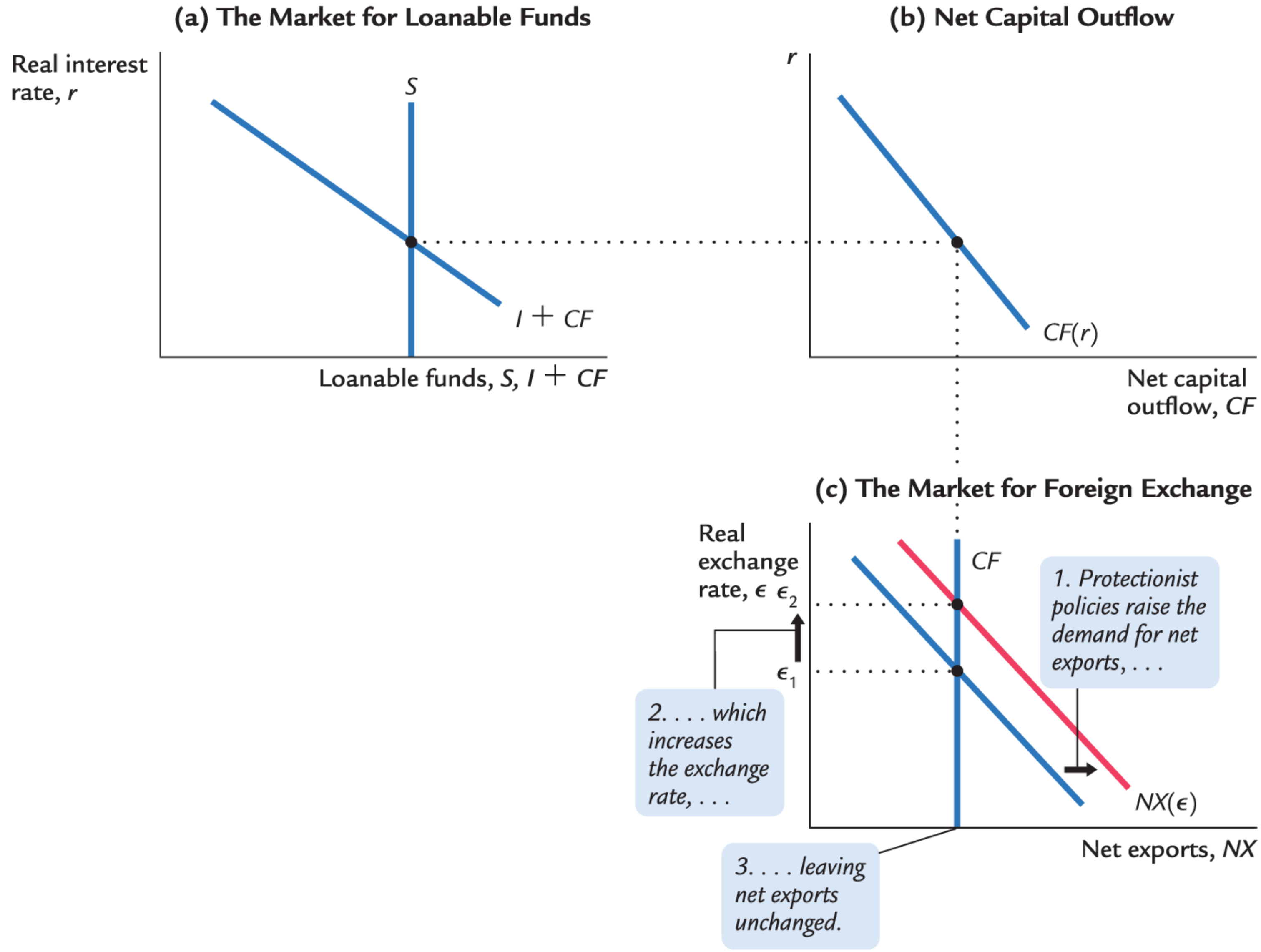


FIGURE 6-22 An Import Restriction in the Large Open Economy An import restriction raises the demand for net exports, as shown in panel (c). The real exchange rate appreciates, while the equilibrium trade balance remains the same. Nothing happens in the market for loanable funds in panel (a) or to the net capital outflow in panel (b).

Shifts in Net Capital Outflow

There are various reasons that the *CF* schedule might shift. One reason is fiscal policy abroad. For example, suppose that Germany pursues a fiscal policy that raises German saving. This policy reduces the German interest rate. The lower German interest rate discourages American investors from lending in Germany and encourages German investors to lend in the United States. For any given U.S. interest rate, the U.S. net capital outflow falls.

Another reason the *CF* schedule might shift is political instability abroad. Suppose that a war or revolution breaks out in another country. Investors around the world will try to withdraw their assets from that country and seek a "safe haven" in a stable country such as the United States. The result is a reduction in the U.S. net capital outflow.

Figure 6-23 shows the impact of a leftward shift in the *CF* schedule. The reduced demand for loanable funds lowers the equilibrium interest rate. The lower interest rate tends to raise net capital outflow, but because this only partly mitigates the shift in the *CF* schedule, *CF* still falls. The reduced level of net capital outflow reduces the supply of dollars in the market for foreign exchange. The exchange rate appreciates, and net exports fall.

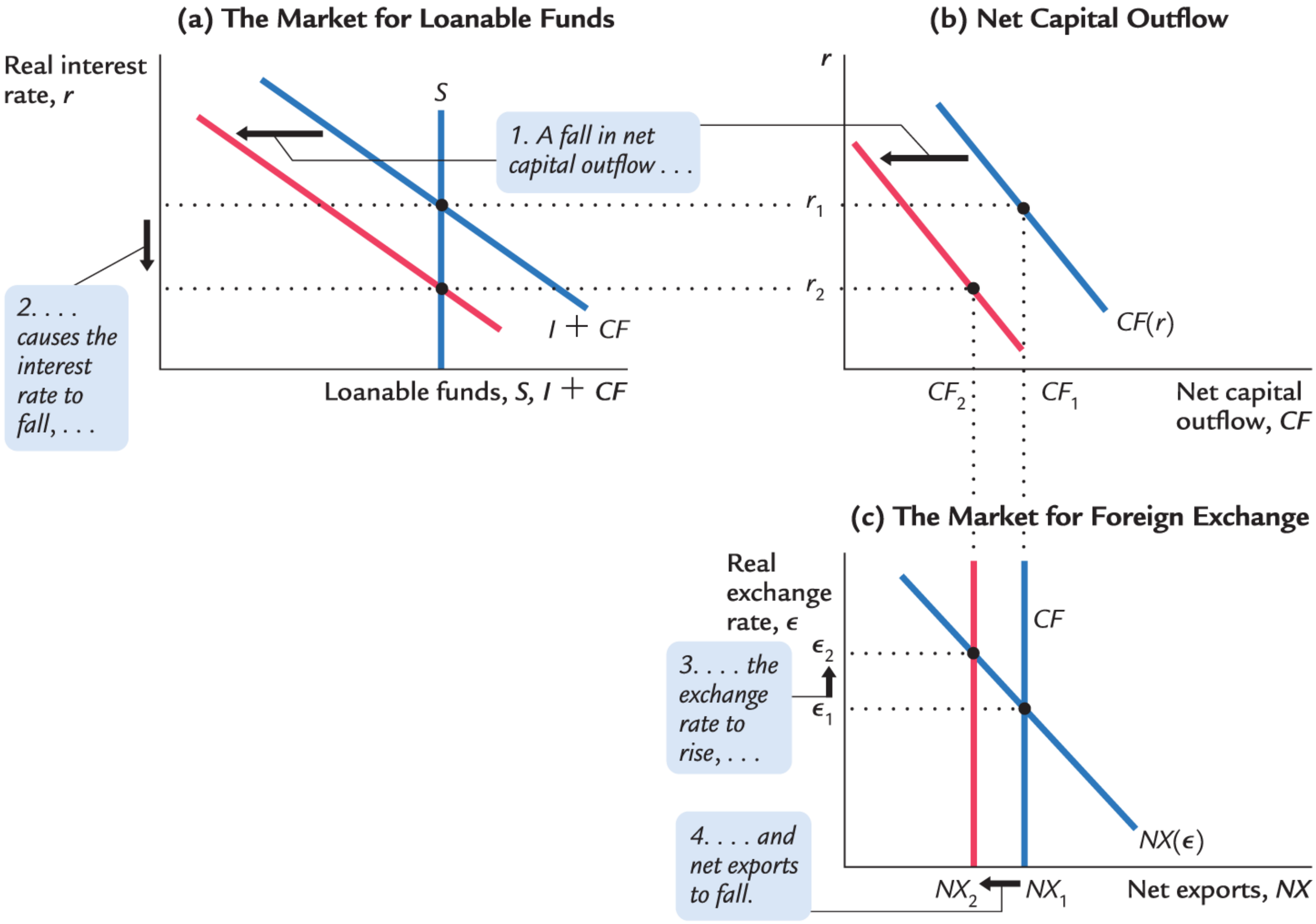


FIGURE 6-23 A Fall in the Net Capital Outflow in the Large Open Economy Panel (a) shows that a downward shift in the *CF* schedule reduces the demand for loans and thereby reduces the equilibrium interest rate. Panel (b) shows that the level of the net capital outflow falls. Panel (c) shows that the real exchange rate appreciates and net exports fall.

Conclusion

How different are large and small open economies? Certainly, policies affect the interest rate in a large open economy, unlike in a small open economy. But, in other ways, the two models yield similar conclusions. In both large and small open economies, policies that raise saving or lower investment lead to trade surpluses. Similarly, policies that lower saving or raise investment lead to trade deficits. In both economies, protectionist trade policies cause the exchange rate to appreciate and do not influence the trade balance. Because the results are so similar, for most questions one can use the simpler model of the small open economy, even if the economy being examined is not really small.

MORE PROBLEMS AND APPLICATIONS

1. If a war broke out abroad, it would affect the U.S. economy in many ways. Use the model of the large open economy to examine each of the following effects of such a war. What happens in the United States to saving, investment, the trade balance, the interest rate, and the exchange rate? (To keep things simple, consider each of the following effects separately.)

    a. The U.S. government, fearing it may need to enter the war, increases its purchases of military equipment.

    b. Other countries raise their demand for high-tech weapons, a major export of the United States.

    c. The war makes U.S. firms uncertain about the future, and the firms delay some investment projects.

    d. The war makes U.S. consumers uncertain about the future, and the consumers save more in response.

    e. Americans become apprehensive about traveling abroad, so more of them spend their vacations in the United States.

    f. Foreign investors seek a safe haven for their portfolios in the United States.

2. On September 21, 1995, “House Speaker Newt Gingrich threatened today to send the United States into default on its debt for the first time in the nation’s history, to force the Clinton Administration to balance the budget on Republican terms” (*New York Times*, September 22, 1995, p. A1). That same day, the interest rate on 30-year U.S. government bonds rose from 6.46 to 6.55 percent, and the dollar fell in value from 102.7 to 99.0 yen. Use the model of the large open economy to explain this event.

CHAPTER 7

Unemployment and the Labor Market

> A man willing to work, and unable to find work, is perhaps the saddest sight that fortune's inequality exhibits under this sun.
>
> —Thomas Carlyle

Unemployment is the macroeconomic problem that affects people most directly and severely. For most people, the loss of a job means a reduced living standard and psychological distress. It is no surprise that unemployment is often a topic of political debate, with politicians claiming their proposed policies would help create jobs.

Economists study unemployment to identify its causes and to help improve the public policies that affect the unemployed. Some of these policies, such as job-training programs, help people find employment. Others, such as unemployment insurance, alleviate the hardships that the unemployed face. Still other policies affect the prevalence of unemployment inadvertently. Laws mandating a high minimum wage, for instance, are thought to raise unemployment among the least skilled and least experienced members of the labor force.

Our discussions of the labor market so far have ignored unemployment. In particular, the model of national income in Chapter 3 was built with the assumption that the economy is always at full employment. In reality, however, not everyone in the labor force has a job all the time: in all free-market economies, at any moment, some people are unemployed.

Figure 7-1 shows the rate of unemployment—the percentage of the labor force unemployed—in the United States from 1950 to 2017. Although the rate of unemployment fluctuates from year to year, it never gets close to zero. The average is between 5 and 6 percent, meaning that out of every eighteen people wanting a job, one person does not have one.

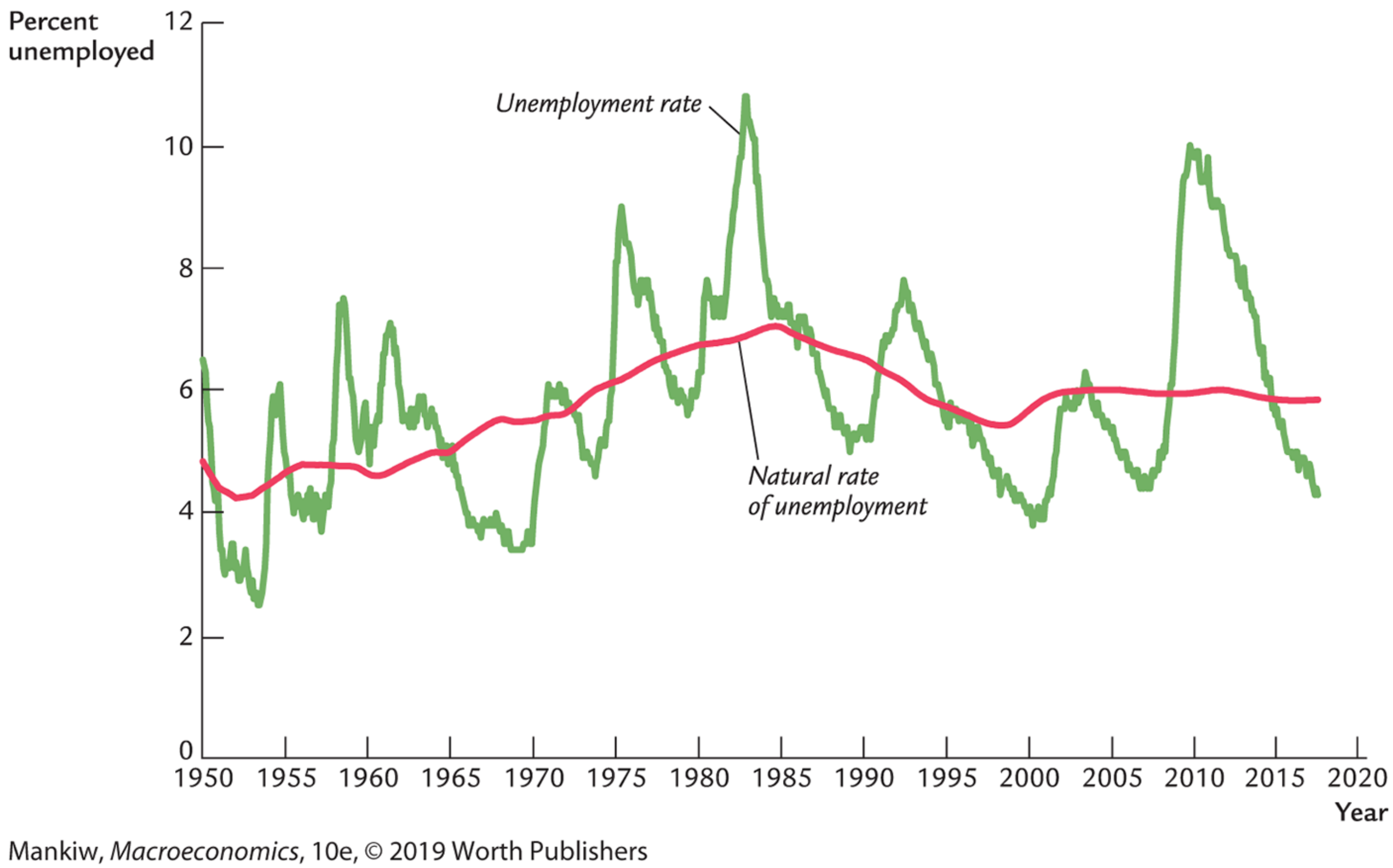


FIGURE 7-1 The Unemployment Rate and the Natural Rate of Unemployment in the United States There is always some unemployment. The natural rate of unemployment is the average level around which the unemployment rate fluctuates. (The natural rate of unemployment for any particular month is estimated here by averaging all the unemployment rates from ten years earlier to ten years later. Future unemployment rates are set at 5.5 percent.)

Data from: Bureau of Labor Statistics.

In this chapter we begin our study of unemployment by discussing why there is always some unemployment and what determines its level. We do not study what determines the year-to-year fluctuations in the rate of unemployment until Part Four of this book, which examines short-run economic fluctuations. Here we examine the determinants of the **natural rate of unemployment**—the average rate of unemployment around which the economy fluctuates. The natural rate is the rate of unemployment toward which the economy gravitates in the long run, given all the labor-market imperfections that impede workers from instantly finding jobs.

7-1 Job Loss, Job Finding, and the Natural Rate of Unemployment

Every day some workers lose or quit their jobs, and some unemployed workers are hired. This perpetual ebb and flow determines the fraction of the labor force that is unemployed. In this section we develop a model of labor-force dynamics that shows what determines the natural rate of unemployment.[1]

We start with some notation. Let L denote the labor force, E the number of employed workers, and U the number of unemployed workers. Because every worker is either employed or unemployed, the labor force is the sum of the employed and the unemployed:

$$L = E + U.$$

Using this notation, the rate of unemployment is U/L.

To see what determine the unemployment rate, we assume that the labor force L is fixed and focus on the transition of individuals in the labor force between employment E and unemployment U. This is illustrated in Figure 7-2. Let s denote the *rate of job separation*, the fraction of employed individuals who lose or leave their jobs each month. Let f denote the *rate of job finding*, the fraction of unemployed individuals who find a job each month. Together, the rate of job separation s and the rate of job finding f determine the rate of unemployment.

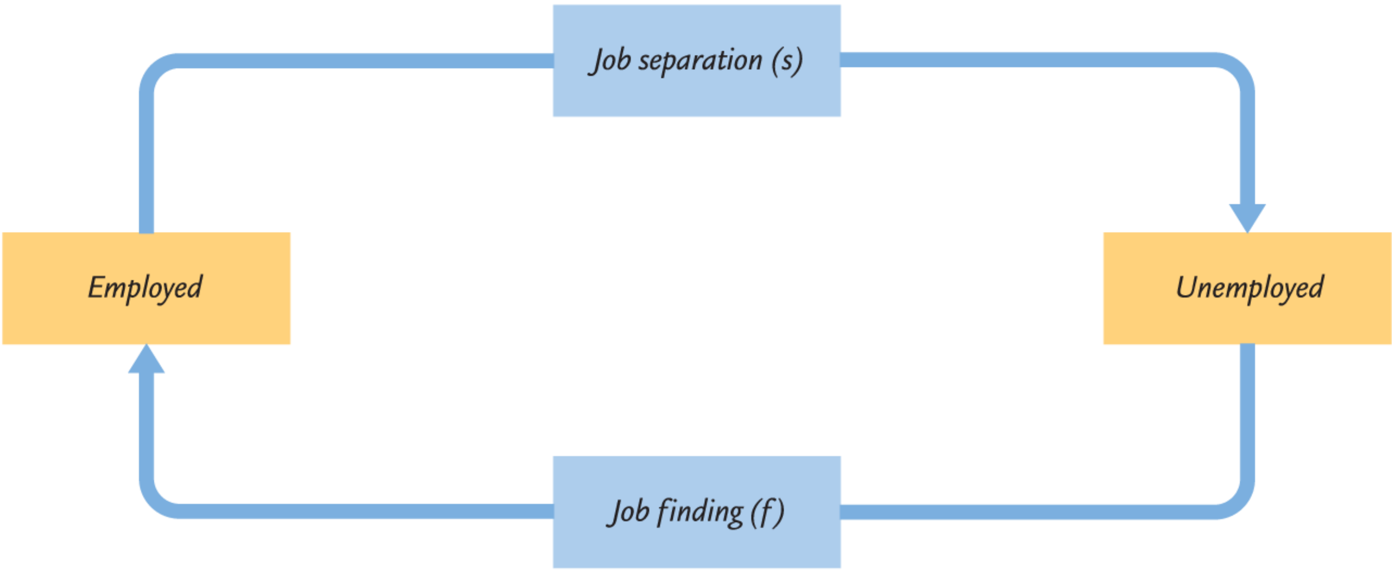


FIGURE 7-2 The Transitions Between Employment and Unemployment In every period, a fraction s of the employed lose their jobs, and a fraction f of the unemployed find jobs. The rates of job separation and job finding determine the

rate of unemployment.

If the unemployment rate is neither rising nor falling—that is, if the labor market is in a *steady state*—then the number of people finding jobs *f U* must equal the number of people losing jobs *sE*. We write the steady-state condition as

$$fU = sE.$$

We can use this equation to find the steady-state unemployment rate. From our definition of the labor force, we know that $E = L - U$; that is, the number of employed equals the labor force minus the number of unemployed. If we substitute $(L - U)$ for *E* in the steady-state condition, we get

$$fU = s(L - U).$$

Next, we divide both sides of this equation by *L* to obtain

$$f\frac{U}{L} = s\left(1 - \frac{U}{L}\right).$$

Now we can solve for the unemployment rate *U/L* to find

$$\frac{U}{L} = \frac{s}{s+f}.$$

This can also be written as

$$\frac{U}{L} = \frac{1}{1+f/s}.$$

This equation shows that the steady-state rate of unemployment *U/L* depends on the rates of job separation *s* and job finding *f*. The higher the rate of job separation, the higher the unemployment rate. The higher the rate of job finding, the lower the unemployment rate.

Here's a numerical example. Suppose that 1 percent of the employed lose their jobs each month $(s = 0.01)$. This means that the average spell of employment lasts 1/0.01, or 100 months, about 8 years.

Suppose further that 20 percent of the unemployed find a job each month (f=0.20), $(f = 0.20)$, so that the average spell of unemployment last 5 months. Then the steady-state rate of unemployment is

UL=0.010.01+0.20=0.0476.

$$\frac{U}{L} = \frac{0.01}{0.01 + 0.20}$$
$$= 0.0476.$$

The rate of unemployment in this example is about 5 percent.

This simple model of the natural rate of unemployment has an important implication for public policy. *Any policy aimed at lowering the natural rate of unemployment must either reduce the rate of job separation or increase the rate of job finding. Similarly, any policy that affects the rate of job separation or job finding also changes the natural rate of unemployment.*

Although this model is useful in relating the unemployment rate to job separation and job finding, it fails to answer a central question: why is there unemployment in the first place? If a person could always find a job quickly, the rate of job finding would be very high and the rate of unemployment would be near zero. This model of the unemployment rate assumes that job finding is not instantaneous, but it fails to explain why. In the next two sections, we examine two reasons for unemployment: job search and wage rigidity.

7-2 Job Search and Frictional Unemployment

One reason for unemployment is that it takes time to match workers and jobs. The equilibrium model of the aggregate labor market discussed in Chapter 3 assumes that all workers and all jobs are identical and, therefore, that all workers are equally well suited to all jobs. If this were true and the labor market were in equilibrium, a job loss would not cause unemployment: a laid-off worker would immediately find a new job at the market wage.

In fact, workers have different preferences and abilities, and jobs have different attributes. Furthermore, the flow of information about job candidates and job vacancies is imperfect, and the geographic mobility of workers is not instantaneous. For all these reasons, searching for a job takes time and effort, and this tends to reduce the rate of job finding. Indeed, because different jobs require different skills and pay different wages, unemployed workers may not accept the first job offer they receive. The unemployment caused by the time it takes workers to search for a job is called **frictional unemployment**.

Causes of Frictional Unemployment

Some frictional unemployment is inevitable in a changing economy. For many reasons, the types of goods that firms and households demand vary over time. As the demand for goods shifts, so does the demand for the labor that produces those goods. The invention of the personal computer, for example, reduced the demand for typewriters and the demand for labor by typewriter manufacturers. At the same time, it increased the demand for labor in the electronics industry. Similarly, because different regions produce different goods, the demand for labor may be rising in one part of the country and falling in another. An increase in the price of oil may cause the demand for labor to rise in oil-producing states such as Texas, but because expensive oil means expensive gasoline, it makes driving less attractive and may decrease the demand for labor in auto-producing states such as Michigan. Economists call a change in the composition of demand among industries or regions a **sectoral shift**. Because sectoral shifts are always occurring, and because it takes time for workers to change sectors, there is always frictional unemployment.

Sectoral shifts are not the only cause of job separation and frictional unemployment. In addition, workers find themselves out of work when their firms fail, when their job performance is deemed unacceptable, or when their particular skills are no longer needed. Workers also may quit their jobs to change careers or to move to different parts of the country. Regardless of the cause of the job separation, it takes time and effort for the worker to find a new job. As long as the supply and demand for labor among firms is changing, frictional unemployment is unavoidable.

Public Policy and Frictional Unemployment

Many public policies seek to decrease the natural rate of unemployment by reducing frictional unemployment. Government employment agencies disseminate information about job vacancies to match jobs and workers more efficiently. Publicly funded retraining programs are designed to ease the transition of workers from declining to growing industries. If these programs succeed at increasing the rate of job finding, they decrease the natural rate of unemployment.

Other government programs inadvertently increase frictional unemployment. One of these is **unemployment insurance**. Under this program, unemployed workers can collect a fraction of their wages for a certain period after losing their jobs. Although the precise terms of the program differ from year to year and from state to state, a typical worker covered by unemployment insurance in the United States receives 50 percent of her former wages for 26 weeks. In many European countries, unemployment-insurance programs are significantly more generous.

By softening the hardship of unemployment, unemployment insurance increases the amount of frictional unemployment and raises the natural rate. The unemployed who receive unemployment-insurance benefits are less pressed to search for new employment and are more likely to turn down unattractive job offers. Both of these changes in behavior reduce the rate of job finding. In addition, because workers know that their incomes are partially protected by unemployment insurance, they are less likely to seek jobs with stable employment prospects and are less likely to bargain for guarantees of job security. These behavioral changes raise the rate of job separation.

Even though unemployment insurance increases the natural rate of unemployment, we should not infer that the policy is ill advised. The program has the benefit of reducing workers' uncertainty about their incomes. Moreover, inducing workers to reject unattractive job offers may lead to better matches between workers and jobs. Evaluating the costs and benefits of different systems of unemployment insurance is a difficult task that continues to be a topic of research.

Economists often propose reforms to the unemployment-insurance system that would reduce unemployment. One proposal is to require a firm that lays off a worker to bear the full cost of that worker's unemployment benefits. Such a system is called *100 percent experience rated* because the rate that each firm pays into the unemployment-insurance system fully reflects the unemployment experience of its own workers. Most current programs are *partially experience rated*. Under this system, when a firm lays off a worker, it is charged for only part of the worker's unemployment benefits; the remainder comes from the program's general revenue. Because a firm pays only a fraction of the cost of the unemployment it causes, it has an incentive to lay off workers when its demand for labor is temporarily low. By reducing that incentive, the proposed reform may reduce the prevalence of temporary layoffs.

CASE STUDY

Unemployment Insurance and the Rate of Job Finding

Many studies have examined the effect of unemployment insurance on job search. The most persuasive studies use data on the experiences of unemployed individuals rather than economy-wide rates of unemployment. Individual data often yield sharp results open to fewer alternative explanations.

One study followed the experience of workers as they used up their eligibility for unemployment-insurance benefits. It found that when unemployed workers become ineligible for benefits, they are more likely to find jobs. In particular, the probability of a person finding a job more than doubles when her benefits run out. One possible explanation is that an absence of benefits increases the search effort of unemployed workers. Another possibility is that workers without benefits are more likely to accept job offers they would otherwise decline because of low wages or poor working conditions.[2]

Additional evidence on how economic incentives affect job search comes from an experiment that the state of Illinois ran in 1985. Randomly selected new claimants for unemployment insurance were each offered a $500 bonus if they found employment within 11 weeks. The subsequent experience of this group was compared to that of a control group not offered this incentive. The average duration of unemployment for the group offered the $500 bonus was 17.0 weeks, compared to 18.3 weeks for the control group. Thus, the prospect of earning the bonus reduced the average spell of unemployment by 7 percent, suggesting that more effort was devoted to job search. This experiment shows clearly that the incentives provided by the unemployment-insurance system affect the rate of job finding.[3] ■

7-3 Real-Wage Rigidity and Structural Unemployment

A second reason for unemployment is **wage rigidity**—the failure of wages to adjust to a level at which labor supply equals labor demand. In the equilibrium model of the labor market, as outlined in Chapter 3, the real wage adjusts to equilibrate labor supply and labor demand. Yet wages are not always flexible. Sometimes the real wage is stuck above the market-clearing level.

Figure 7-3 shows why wage rigidity leads to unemployment. When the real wage is above the level that equilibrates supply and demand, the quantity of labor supplied exceeds the quantity demanded. Firms must in some way ration the scarce jobs among workers. Real-wage rigidity reduces the rate of job finding and raises the level of unemployment.

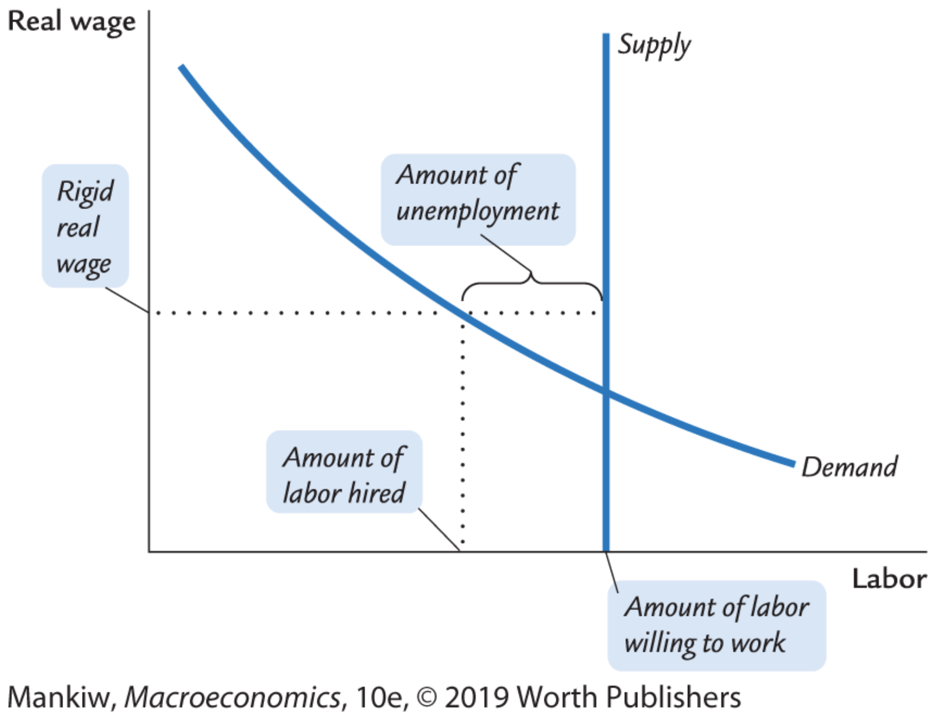


FIGURE 7-3 Real-Wage Rigidity Leads to Job Rationing If the real wage is stuck above the equilibrium level, then the supply of labor exceeds the demand. The result is unemployment.

The unemployment resulting from wage rigidity and job rationing is sometimes called **structural unemployment**. Workers are unemployed not because they are actively searching for the jobs that best suit their skills but because there is a mismatch between the number of people who want to work and the number of jobs that are available. At the going wage, the quantity of labor supplied exceeds the quantity of labor demanded; many workers are simply waiting for jobs to open up.

To understand wage rigidity and structural unemployment, we must examine why the labor market does not clear. When the real wage exceeds the equilibrium level and the supply of workers exceeds the demand, we might expect firms to lower the wages they pay. Structural unemployment arises because firms fail to reduce wages despite an excess supply of labor. We now turn to three causes of this wage rigidity: minimum-wage laws, the monopoly power of unions, and efficiency wages.

Minimum-Wage Laws

The government causes wage rigidity when it prevents wages from falling to equilibrium levels. Minimum-wage laws set a legal minimum on the wages that firms pay their employees. Since the passage of the Fair Labor Standards Act of 1938, the U.S. federal government has enforced a minimum wage that has usually been between 30 and 50 percent of the average wage in manufacturing. In addition, many states and cities enact minimum wages that are higher than the federal one: for example, in 2017, when the federal minimum wage was $7.25 per hour, California had a minimum wage of $10 per hour, and Seattle had a minimum wage of $15 per hour for large employers. For most workers, the minimum wage is not binding because they earn well above the legislated minimum. Yet for some workers, especially the unskilled and inexperienced, the minimum wage raises their wage above its equilibrium level and, therefore, reduces the quantity of their labor that firms demand.

Economists believe that the minimum wage has its greatest impact on teenage unemployment. The equilibrium wages of teenagers tend to be low for two reasons. First, because teenagers are among the least skilled and least experienced members of the labor force, they tend to have low marginal productivity. Second, teenagers often take some "compensation" in the form of on-the-job training rather than direct pay. An internship is a classic example of training offered in place of wages. For both reasons, the wage at which the supply of teenage workers equals the demand is low. The minimum wage is therefore more often binding for teenagers than for others in the labor force. Empirical studies typically find that a 10 percent increase in the minimum wage reduces teenage employment by 1 to 3 percent.[4]

The minimum wage is a perennial source of political debate. Advocates of a higher minimum wage view it as a way to raise the income of the working poor. Certainly, the minimum wage provides only a meager standard of living: in the United States, a single parent with one child working full time at a minimum-wage job would fall below the official poverty level for a family of that size. Although minimum-wage advocates often admit that the policy causes unemployment for some workers, they argue that this cost is worth bearing to raise others out of poverty.

Opponents of a higher minimum wage claim that it is not the best way to help the working poor. They contend not only that the increased labor costs raise unemployment but also that the minimum wage is poorly targeted. Many minimum-wage earners are teenagers from middle-class homes working for discretionary spending money rather than heads of households working to support their families.

One might have hoped that empirical research could close this political divide. Unfortunately, different studies using varying data and methodologies often reach conflicting results. The large increase in the minimum wage in Seattle from 2014 to 2016 is a case in point. One study of the Seattle food services industry concluded that wages increased significantly without a discernible effect on employment.[5] Another study concluded that hours worked in low-wage jobs fell by about 9 percent, while wages increased by only 3 percent, indicating that workers' incomes fell as a result of the minimum wage hike.[6] One drawback of most minimum-wage studies is that they focus on the effects over short periods of time (such as by comparing employment the year before and the year after a change in the minimum wage). The longer-term effects on employment are arguably more relevant for evaluating the policy, but these are harder to estimate.

When judging the minimum wage, it is useful to keep in mind alternative policies. Many economists believe that tax credits are a better way to increase the incomes of the working poor. The *earned income tax credit* is an amount that poor working families are allowed to subtract from the taxes they owe. For a family with very low income, the credit exceeds its taxes, and the family receives a payment from the government. Unlike the minimum wage, the earned income tax credit does not raise labor costs to firms and, therefore, does not reduce the quantity of labor that firms demand. It has the disadvantage, however, of reducing the government's tax revenue.

Unions and Collective Bargaining

A second cause of wage rigidity is the market power of unions. Table 7-1 shows the importance of unions in several major countries. In the United States, only 12 percent of workers have their wages set through collective bargaining. In most European countries, unions play a much larger role.

TABLE 7-1 Percent of Workers Covered by Collective Bargaining

Turkey	7%
South Korea	12
United States	12
Poland	15
Japan	17
Israel	26
Canada	29
United Kingdom	30
Greece	42

Switzerland	49
Germany	58
Australia	60
Spain	78
Italy	80
Netherlands	85
Sweden	89
Belgium	96
France	98
Data from: *Economic Policy Reforms 2017: Going for Growth*, OECD, 2017.	

The wages of unionized workers are determined not by the equilibrium of supply and demand but by bargaining between union leaders and firm management. Often, the final agreement raises the wage above the equilibrium level and allows the firm to decide how many workers to employ. The result is a reduction in the number of workers hired, a lower rate of job finding, and an increase in structural unemployment.

Unions can also influence the wages paid by firms whose workforces are not unionized because the threat of unionization can keep wages above the equilibrium level. Most firms dislike unions. Unions not only raise wages but also increase the bargaining power of labor on many other issues, such as hours of employment and working conditions. A firm may choose to pay its workers high wages to keep them happy and discourage them from forming a union.

The unemployment caused by unions and by the threat of unionization is an instance of conflict between different groups of workers—**insiders** and **outsiders**. Workers already employed by a firm, the insiders, typically try to keep their firm’s wages high. The unemployed, the outsiders, bear part of the cost of higher wages because at a lower wage, they might be hired. These two groups have conflicting interests. The effect of any bargaining process on wages and employment depends on the relative influence of each group.

The conflict between insiders and outsiders is resolved differently in different countries. In some countries, such as the United States, wage bargaining takes place at the level of the firm or plant. In other countries, such as Sweden, wage bargaining takes place at the national level—with the government often playing a key role. Despite a highly unionized labor force, Sweden has not experienced extraordinarily high unemployment throughout its history. One possible explanation is that the centralization of wage bargaining and the role of the government in the bargaining process give more influence to the outsiders, which keeps wages closer to the equilibrium level.

Efficiency Wages

Efficiency-wage theories propose a third cause of wage rigidity in addition to minimum-wage laws and unionization. These theories hold that high wages make workers more productive. The influence of wages on worker efficiency may explain the failure of firms to cut wages despite an excess supply of labor. Even though a wage reduction would lower a firm's wage bill, it would also—if these theories are correct—lower worker productivity and the firm's profits.

Economists have proposed various theories to explain how wages affect worker productivity. One efficiency-wage theory, applied mostly to poorer countries, holds that wages influence nutrition. Better-paid workers can afford a more nutritious diet, and healthier workers are more productive. A firm may decide to pay a wage above the equilibrium level to maintain a healthy workforce. This consideration is not important for employers in wealthier countries, such as the United States and most of Europe, because the equilibrium wage is well above the level necessary to maintain good health.

A second efficiency-wage theory, more relevant for developed countries, holds that high wages reduce labor turnover. Workers quit jobs for many reasons—to accept better positions at other firms, to change careers, or to move to other parts of the country. The more a firm pays its workers, the greater is their incentive to stay with the firm. By paying a high wage, a firm reduces the frequency at which its workers quit, thereby decreasing the time and money spent hiring and training new workers.

A third efficiency-wage theory holds that the quality of a firm's workforce depends on the wage it pays its employees. If a firm reduces its wage, the best employees may take jobs elsewhere, leaving the firm with inferior employees who have fewer alternative opportunities. Economists recognize this unfavorable sorting as an example of *adverse selection*—the tendency of people with more information (in this case, the workers, who know their own outside opportunities) to self-select in a way that disadvantages people with less information (the firm). By paying an above-equilibrium wage, the firm may reduce adverse selection, improve the quality of its workforce, and thereby increase productivity.

A fourth efficiency-wage theory holds that a high wage improves worker effort. This theory posits that firms cannot perfectly monitor their employees' work effort and that employees must themselves decide how hard to work. Workers can work hard, or they can shirk and risk getting caught and fired. Economists recognize this possibility as an example of *moral hazard*—the tendency of people to behave inappropriately when their behavior is imperfectly monitored. The firm can reduce moral hazard by paying a high wage. The higher the wage, the greater the cost to the worker of getting fired. By paying a higher wage, a firm induces more of its employees not to shirk and thus increases their productivity.

These four efficiency-wage theories differ in detail, but they share a common theme: because a firm operates more efficiently if it pays its workers a high wage, the firm may find it profitable to keep wages above the level that balances supply and demand. The result of this higher-than-equilibrium wage is a lower rate of job finding and greater unemployment.[7]

CASE STUDY

Henry Ford's $5 Workday

In 1914 the Ford Motor Company started paying its workers $5 per day. The prevailing wage at the time was between $2 and $3 per day, so Ford's wage was well above the equilibrium level. Not surprisingly, long lines of job seekers waited outside the Ford plant gates, hoping for a chance to earn this high wage.

What was Ford's motive? Henry Ford later wrote, "We wanted to pay these wages so that the business would be on a lasting foundation. We were building for the future. A low wage business is always insecure. . . . The payment of five dollars a day for an eight hour day was one of the finest cost-cutting moves we ever made."

From the standpoint of traditional economic theory, Ford's explanation seems peculiar. He was suggesting that *high* wages imply *low* costs. But perhaps Ford had discovered efficiency-wage theory. Perhaps he was using the high wage to increase worker productivity.

Evidence suggests that paying such a high wage did benefit the company. According to an engineering report written at the time, "The Ford high wage does away with all this inertia and living force resistance. . . . The workmen are absolutely docile, and it is safe to say that since the last day of 1912, every single day has seen marked reductions made in the Ford shops' labor costs." Absenteeism fell by 75 percent, suggesting a large increase in worker effort. Alan Nevins, a historian who studied the early Ford Motor Company, wrote, "Ford and his associates freely declared on many occasions that the high wage policy had turned out to be good business. By this, they meant that it had improved the discipline of the workers, given them a more loyal interest in the institution, and raised their personal efficiency."[8] ■

7-4 Labor-Market Experience: The United States

So far we have developed the theory behind the natural rate of unemployment. We began by showing that the economy's steady-state unemployment rate depends on the rates of job separation and job finding. Then we discussed two reasons job finding is not instantaneous: the process of job search (which leads to frictional unemployment) and wage rigidity (which leads to structural unemployment). Wage rigidity, in turn, arises from minimum-wage laws, unionization, and efficiency wages.

With these theories as background, we now examine some additional facts about unemployment, focusing at first on the case of American labor markets. These facts will help us to evaluate our theories and assess public policies aimed at reducing unemployment.

The Duration of Unemployment

When a person becomes unemployed, is the spell of unemployment likely to be short or long? The answer to this question is important because it indicates the reasons for the unemployment and what policy response is appropriate. On the one hand, if most unemployment is short-term, one might argue that it is frictional and perhaps unavoidable. Unemployed workers may need some time to search for the job that is best suited to their skills and tastes. On the other hand, long-term unemployment cannot easily be attributed to the time it takes to match jobs and workers: we would not expect this matching process to take many months. Long-term unemployment is more likely to be structural unemployment, representing a mismatch between the number of jobs available and the number of people who want to work. Thus, data on the duration of unemployment can affect our view about the reasons for unemployment.

The answer to our question is subtle. *The data show that most spells of unemployment are short but that most weeks of unemployment are attributable to the long-term unemployed.* To see how both these facts can be true, consider an extreme but simple example. Suppose that 10 people are unemployed for part of a given year. Of these 10 people, 8 are unemployed for 1 month and 2 are unemployed for 12 months, totaling 32 months of unemployment. In this example, most spells of unemployment are short: 8 of the 10 unemployment spells, or 80 percent, end in 1 month. Yet most months of unemployment are attributable to the long-term unemployed: 24 of the 32 months of unemployment, or 75 percent, are experienced by the 2 workers who are each unemployed for 12 months. Depending on whether we look at spells of unemployment or months of unemployment, most unemployment can appear to be either short-term or long-term.

This evidence on the duration of unemployment has an important implication for public policy. If the goal is to substantially lower the natural rate of unemployment, policies must aim at the long-term unemployed because these individuals account for a large amount of unemployment. Yet policies must be carefully targeted because the long-term unemployed constitute a small minority of those who become unemployed. Most people who become unemployed find work within a short time.

CASE STUDY

The Increase in U.S. Long-Term Unemployment and the Debate over Unemployment Insurance

In 2008 and 2009, as the U.S. economy experienced a deep recession, the labor market demonstrated a new and striking phenomenon: a large upward spike in the duration of unemployment. Figure 7-4 shows the median duration of unemployment for jobless workers from 1967 to 2017. Recessions are indicated by shaded areas. The figure shows that the duration of unemployment typically rises during recessions. The huge increase during the recession of 2008–2009, however, is without precedent in modern history.

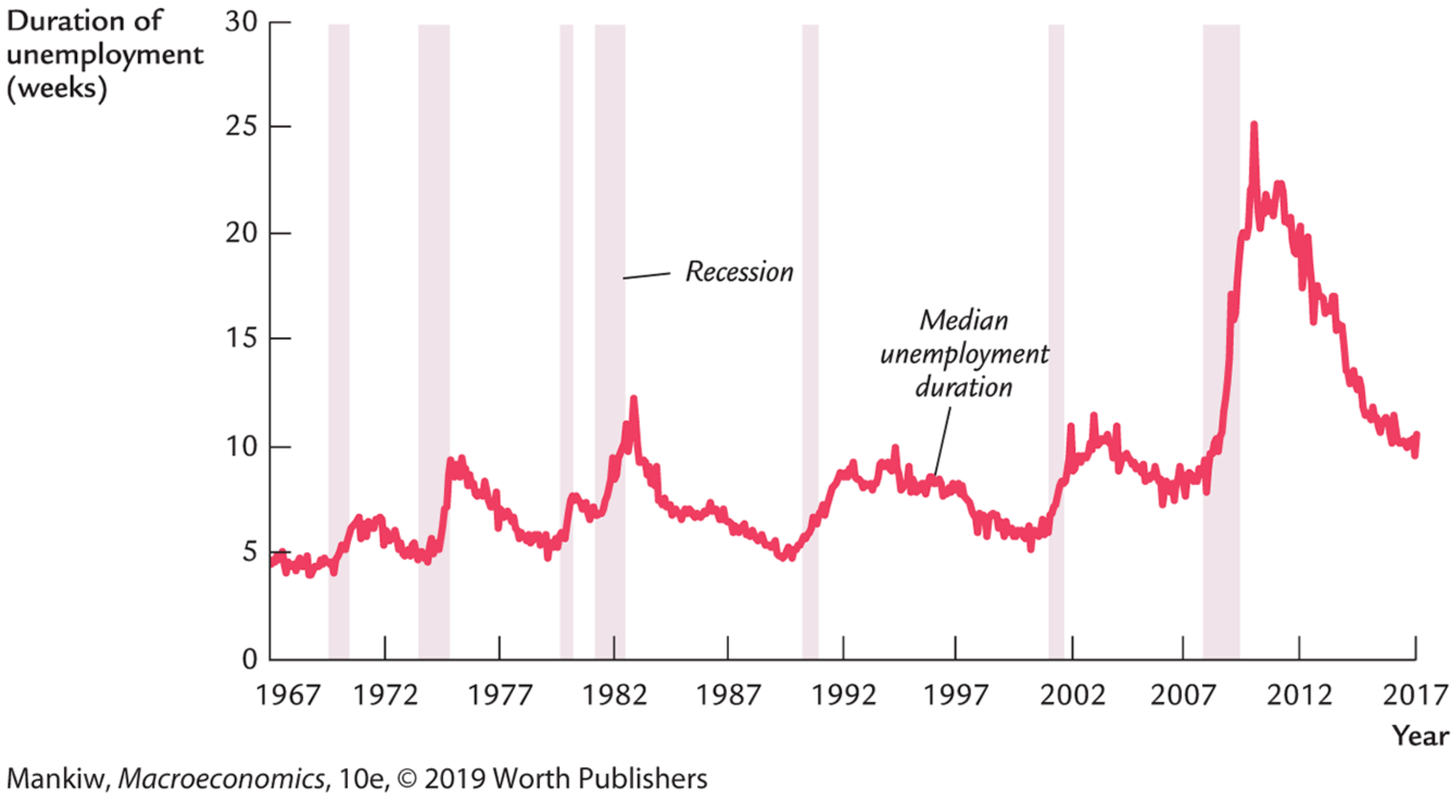


FIGURE 7-4 The Median Duration of Unemployment The median duration of unemployment typically rises during recessions, shown as the shaded areas here, but its spike upward during the recession of 2008–2009 was unprecedented.

Data from: Bureau of Labor Statistics.

What explains this phenomenon? Economists fall into two camps.

Some economists believe that the increase in long-term unemployment is a result of government policies. In February 2009, when the depth of the recession was apparent, Congress extended the eligibility for unemployment insurance from the normal 26 weeks to 99 weeks, and it did not allow this program of extended benefits to expire until January 2014. Extending unemployment-insurance benefits is typical during recessions because jobs are harder to find, but the extension to nearly two years was extraordinary.

Economist Robert Barro wrote an article in the August 30, 2010, issue of the *Wall Street Journal* titled "The

Folly of Subsidizing Unemployment." According to Barro, "the dramatic expansion of unemployment insurance eligibility to 99 weeks is almost surely the culprit" responsible for the rise in long-term unemployment. He writes:

> Generous unemployment insurance programs have been found to raise unemployment in many Western European countries in which unemployment rates have been far higher than the current U.S. rate. In Europe, the influence has worked particularly through increases in long-term unemployment.

Barro concludes that the "reckless expansion of unemployment-insurance coverage to 99 weeks was unwise economically and politically."

Other economists, however, are skeptical that these government policies are to blame. In their view, the increase in eligibility for unemployment insurance was a reasonable and compassionate response to a historically deep downturn and weak labor market.

Here is economist Paul Krugman, writing in a July 4, 2010, *New York Times* article titled "Punishing the Jobless":

> Do unemployment benefits reduce the incentive to seek work? Yes: workers receiving unemployment benefits aren't quite as desperate as workers without benefits, and are likely to be slightly more choosy about accepting new jobs. The operative word here is "slightly": recent economic research suggests that the effect of unemployment benefits on worker behavior is much weaker than was previously believed. Still, it's a real effect when the economy is doing well.
>
> But it's an effect that is completely irrelevant to our current situation. When the economy is booming, and lack of sufficient willing workers is limiting growth, generous unemployment benefits may keep employment lower than it would have been otherwise. But as you may have noticed, right now the economy isn't booming—there are five unemployed workers for every job opening. Cutting off benefits to the unemployed will make them even more desperate for work—but they can't take jobs that aren't there.
>
> Wait: there's more. One main reason there aren't enough jobs right now is weak consumer demand. Helping the unemployed, by putting money in the pockets of people who badly need it, helps support consumer spending.

Barro and Krugman, both prominent economists, have opposite views about this policy debate. The cause of the spike in U.S. long-term unemployment remains an open question. ■

Variation in the Unemployment Rate Across Demographic Groups

The rate of unemployment varies substantially across groups within the population. Table 7-2 presents the U.S.

unemployment rates for various demographic groups in 2016, when the overall unemployment rate was 4.9 percent.

Table 7-2 Unemployment Rate by Demographic Group, 2016

Age	White Men	White Women	Black Men	Black Women
16–19	14.9	13.2	30.9	22.8
20–24	8.0	6.3	17.0	12.3
25–54	3.6	3.7	7.3	7.0

Data from: Bureau of Labor Statistics.

This table shows that younger workers have much higher unemployment rates than older ones. To explain this difference, recall our model of the natural rate of unemployment. The model isolates two possible causes for a high rate of unemployment: a low rate of job finding and a high rate of job separation. When economists study data on the transition of individuals between employment and unemployment, they find that those groups with high unemployment tend to have high rates of job separation. They find less variation across groups in the rate of job finding. For example, an employed white male is four times more likely to become unemployed if he is a teenager than if he is middle aged; once he is unemployed, his rate of job finding is not closely related to his age.

These findings help explain the higher unemployment rates for younger workers. Younger workers have only recently entered the labor market, and they are often uncertain about their career plans. It may be best for them to try different types of jobs before making a long-term commitment to an occupation. We should therefore expect a higher rate of job separation and higher frictional unemployment for this group.

Another fact that stands out from Table 7-2 is that unemployment rates are much higher for blacks than for whites. This phenomenon is not well understood. Data on transitions between employment and unemployment show that the higher unemployment rates for blacks, especially for black teenagers, arise because of both higher rates of job separation and lower rates of job finding. Possible reasons for the lower rates of job finding include less access to informal job-finding networks and discrimination by employers.

Transitions into and out of the Labor Force

So far we have ignored an important aspect of labor-market dynamics: the movement of individuals into and out of the labor force. Our model of the natural rate of unemployment assumes that the labor force is fixed. In this case, the sole reason for unemployment is job separation, and the sole reason for leaving unemployment is job finding.

In fact, movements into and out of the labor force are important. About one-third of the unemployed have only recently entered the labor force. Some of these entrants are young workers still looking for their first jobs; others have worked before but had temporarily left the labor force. In addition, not all unemployment ends with job finding: almost half of all spells of unemployment end in the unemployed person’s withdrawal from the labor market.

Individuals entering and leaving the labor force make unemployment statistics more difficult to interpret. On the one hand, some individuals calling themselves unemployed may not be seriously looking for jobs and perhaps should best be viewed as out of the labor force. Their “unemployment” may not represent a social problem. On the other hand, some individuals may want jobs but, after unsuccessful searches, have given up looking. These **discouraged workers** are counted as being out of the labor force and do not show up in unemployment statistics. Even though their joblessness is unmeasured, it may nonetheless be a social problem.

Because of these and many other issues that complicate the interpretation of the unemployment data, the Bureau of Labor Statistics calculates several measures of labor underutilization. Table 7-3 gives the definitions and their values as of July 2017. The measures range from 1.7 to 8.6 percent, depending on the characteristics used to classify a worker as not fully employed.

Table 7-3 Alternative Measures of Labor Underutilization

Variable	Description	Rate
U-1	Persons unemployed 15 weeks or longer, as a percent of the civilian labor force (includes only very long-term unemployed)	1.7%
U-2	Job losers and persons who have completed temporary jobs, as a percent of the civilian labor force (excludes job leavers)	2.1
U-3	Total unemployed, as a percent of the civilian labor force (official unemployment rate)	4.3
U-4	Total unemployed, plus discouraged workers, as a percent of the civilian labor force plus discouraged workers	4.7
U-5	Total unemployed plus all marginally attached workers, as a percent of the civilian labor force plus all marginally attached workers	5.3
U-6	Total unemployed, plus all marginally attached workers, plus total employed part time for economic reasons, as a percent of the civilian labor force plus all marginally attached workers	8.6

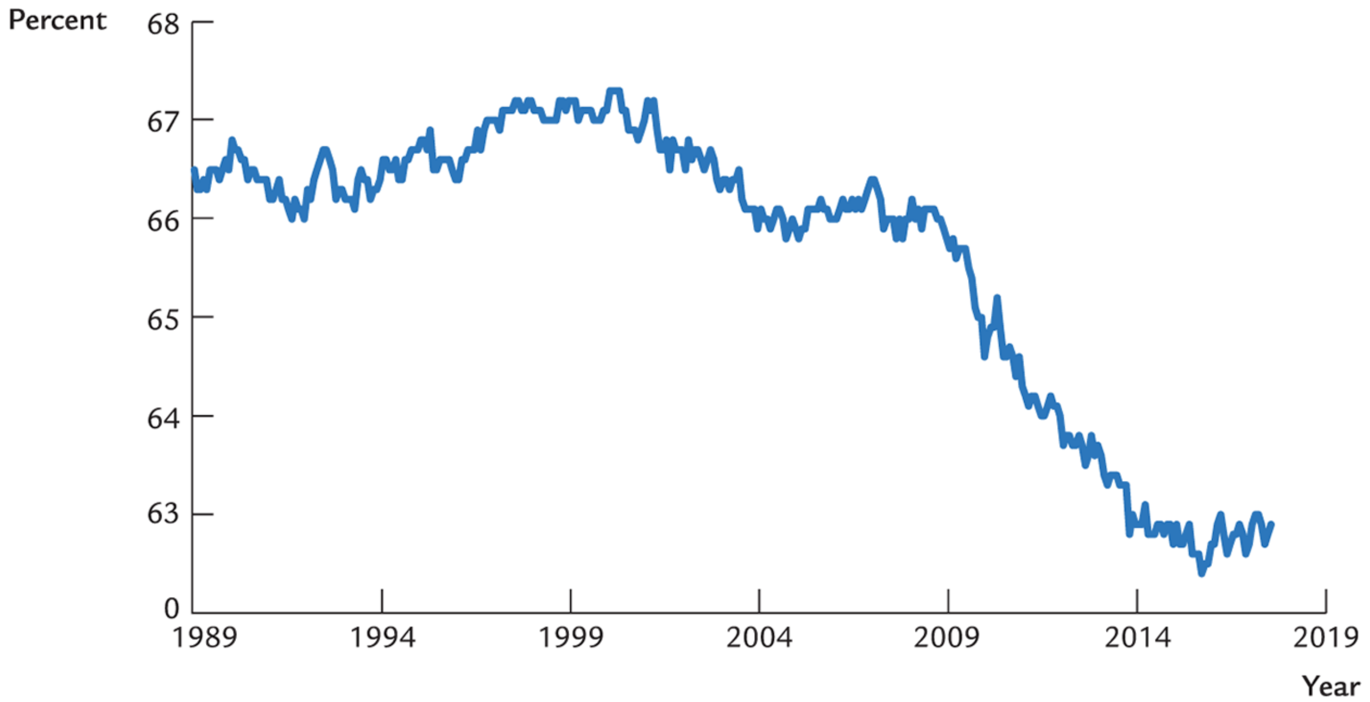
Note: *Marginally attached workers* are persons who currently are neither working nor looking for work but indicate that they want and are available for a job and have looked for work sometime in the recent past. *Discouraged workers*, a subset of the marginally attached, have given a job-market-related reason for not currently looking for a job. *Persons employed part time for economic reasons* are those who want and are available for full-time work but have had to settle for a part-time schedule.

Data from: U.S. Department of Labor. Data are for July 2017.

CASE STUDY

The Decline in Labor-Force Participation: 2007 to 2017

One striking recent development in the U.S. labor market is the decline in labor-force participation. Figure 7-5 illustrates the phenomenon. From 1990 to 2007, the labor-force participation rate fluctuated in a narrow range between about 66 and 67 percent. But then it started a gradual but significant decline. From the fourth quarter of 2007 to the first quarter of 2017, the labor-force participation rate fell from 66.1 percent to 62.8 percent. As a result, about 8 million fewer Americans were working or looking for work in 2017 than otherwise would have been the case.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 7-5 The Labor-Force Participation Rate The labor-force participation rate declined significantly from 2007 to 2014.

Data from: Bureau of Labor Statistics.

What explains the decline of 3.3 percentage points in the labor-force participation rate? To answer this question, a natural place to start is to study those individuals not in the labor force to see why they are not working or looking for work. Economist Shigeru Fujita of the Federal Reserve Bank of Philadelphia has done exactly that, using the data in the Current Population Survey.

His findings, summarized in Table 7-4, allocate the 3.3 percentage points among five categories:

- An increase in retired workers accounts for 2.2 percentage points.
- An increase in disabled workers accounts for 0.6 percentage points.
- An increase in discouraged workers accounts for 0.1 percentage points.
- An increase in those not wanting a job because they are in school accounts for 0.3 percentage points.
- The “other” category—those outside the labor force who are not retired, disabled, discouraged, or in school, such as full-time parents—accounts for none of the change. In fact, this last category went slightly in the other direction.

TABLE 7-4 Decomposing the Change in Labor-Force Participation

		------------------------ By Reason for Nonparticipation ------------------------				
Quarter	Nonparticipation	Retired	Disabled	Discouraged	In School	Other
2007: Q4	33.9%	15.5%	4.9%	1.9%	4.6%	7.1%

2017: Q1	37.2	17.7	5.5	2.0	4.9	6.9
Change	+ 3.3	+ 2.2	+ 0.6	+ 0.1	+ 0.3	− 0.2

Data from: Shigeru Fugita, "On the Causes of Declines in the Labor Force Participation Rate," Federal Reserve Bank of Philadelphia, 2014, and online updates. Parts may not add to total because of rounding.

According to the numbers in Table 7-4, retirement explains the largest share of the increase in nonparticipation, accounting for two-thirds of the change. The increase in the number of retired workers is largely due to the aging of the large baby-boom generation. The baby boom started in 1946, just after World War II, as soldiers returned home and started families, and continued until 1964. The first of the baby boomers turned 62—the earliest age at which a person can start collecting Social Security retirement benefits—in 2008. So far, we have seen just the tip of a sizable iceberg: as more baby boomers reach retirement age in the years to come, the labor-force participation rate will probably continue to decline.

To be sure, this development is not entirely adverse. Retirement is often a welcome change in lifestyle after a lifetime of toil. Yet the decline in labor-force participation does have a cost. With a smaller labor force, the economy produces a smaller output of goods and services, which means a lower level of real GDP.

Moreover, the aging of the population is not the only force at work. Labor-force participation among those in their prime working years (ages 25 to 54) fell from 82.9 percent in the fourth quarter of 2007 to 81.6 percent in the first quarter of 2017. This decline of 1.3 percentage points is smaller than for the overall population but is harder to explain. Proposed hypotheses include (1) increased reliance on government disability programs, (2) weak demand for unskilled workers due to globalization and skilled-biased technological change, (3) higher rates of addiction to opioid drugs, and (4) expanded availability of video games, which increases the value of leisure. All four of these forces may be at work. ■

7-5 Labor-Market Experience: Europe

Although our discussion has focused on the United States, many fascinating and sometimes puzzling phenomena become apparent when economists compare the experiences of Americans in the labor market with those of Europeans.

The Rise in European Unemployment

Figure 7-6 shows the rate of unemployment from 1960 to 2014 in the four largest European countries—France, Germany, Italy, and the United Kingdom. As you can see, the rates of unemployment in these countries have risen substantially. For France, for example, unemployment averaged below 2 percent in the 1960s and above 8 percent in recent years.

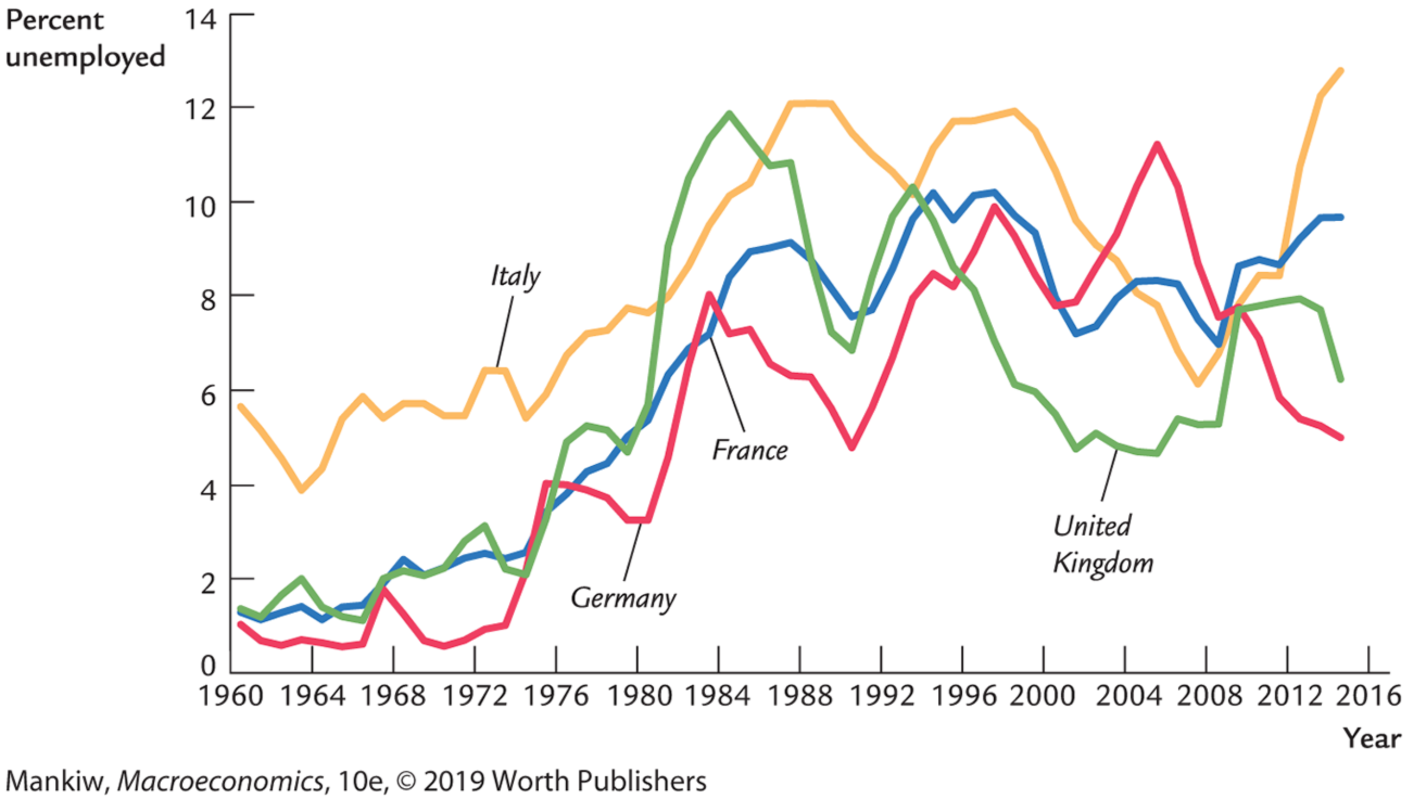


FIGURE 7-6 Unemployment in Europe This figure shows the unemployment rates in the four largest nations in Europe. The figure shows that the European unemployment rate has risen substantially over time, especially in France and Germany.

Data from: OECD.

What is the cause of rising European unemployment? No one knows for sure, but there is a leading theory. Many economists believe that the problem can be traced to the interaction between a long-standing policy and a more recent shock. The long-standing policy is generous benefits for unemployed workers. The recent shock is a technologically driven fall in the demand for unskilled workers relative to skilled workers.

There is no question that most European countries have generous programs for those without jobs. These programs go by various names: social insurance, the welfare state, or simply “the dole.” Many countries allow the unemployed to collect benefits for years rather than for only a short period of time, as in the United States. In some sense, those living on the dole are really out of the labor force: given the employment opportunities available, taking a job is less attractive than remaining without work. Yet these people are often counted as unemployed in government statistics.

There is also no question that the demand for unskilled workers has fallen relative to the demand for skilled workers. This change in demand is probably due to changes in technology: computers, for example, increase the demand for workers who can use them and reduce the demand for those who cannot. In the United States, this change in demand has been reflected in wages rather than unemployment: over the past three decades, the wages of unskilled workers have fallen substantially relative to the wages of skilled workers. In Europe, however, the welfare state provides unskilled workers with an alternative to working for low wages. As the wages of unskilled workers fall, more workers view the dole as their best available option. The result is higher unemployment.

This diagnosis of high European unemployment does not suggest an easy remedy. Reducing the magnitude of government benefits for the unemployed would encourage workers to get off the dole and accept low-wage jobs. But it would also exacerbate economic inequality—the very problem that welfare-state policies were designed to address.[9]

Unemployment Variation Within Europe

Europe is not a single labor market but is, instead, a collection of national labor markets, separated not only by national borders but also by differences in culture and language. Because these countries differ in their labor-market policies and institutions, variation within Europe provides a useful perspective on the causes of unemployment. Many empirical studies have, therefore, focused on these international differences.

The first noteworthy fact is that the unemployment rate varies substantially from country to country. For example, in June 2017, when the unemployment rate was 4.4 percent in the United States, it was 3.8 percent in Germany and 17.1 percent in Spain. Although in recent years average unemployment has been higher in Europe than in the United States, many Europeans live in nations with unemployment rates lower than the U.S. rate.

A second notable fact is that much of the variation in unemployment rates is attributable to the long-term unemployed. The unemployment rate can be separated into two pieces—the percentage of the labor force that has been unemployed for less than a year and the percentage of the labor force that has been unemployed for more than a year. The long-term unemployment rate exhibits more variability from country to country than does the short-term unemployment rate.

National unemployment rates are correlated with various labor-market policies. Unemployment rates are higher in nations with more generous unemployment insurance, as measured by the replacement rate—the percentage of previous wages that is replaced when a worker loses a job. In addition, nations tend to have higher unemployment, especially higher long-term unemployment, if benefits can be collected for longer periods of time.

Although government spending on unemployment insurance seems to raise unemployment, spending on "active" labor-market policies appears to decrease it. These active labor-market policies include job training, assistance with job search, and subsidized employment. Spain, for instance, has historically had a high rate of unemployment, a fact that can be explained by the combination of generous payments to the unemployed with minimal assistance at helping them find new jobs.

The role of unions also varies from country to country, as we saw in Table 7-1. This fact also helps explain differences in labor-market outcomes. National unemployment rates are positively correlated with the percentage of the labor force whose wages are set by collective bargaining with unions. The adverse impact of unions on unemployment is smaller, however, in nations where there is substantial coordination among employers in bargaining with unions, perhaps because coordination may moderate the upward pressure on wages.

A word of warning: correlation does not imply causation, so empirical results such as these should be interpreted with caution. But they do suggest that a nation's unemployment rate, rather than being immutable, is instead a function of the choices a nation makes.[10]

The Rise of European Leisure

Higher unemployment rates in Europe are part of the larger phenomenon that Europeans typically work fewer hours than do their American counterparts. Figure 7-7 shows how many hours a typical person works in the United States, France, and Germany. In the early 1970s, the number of hours worked was about the same in each of these countries. But since then, the number of hours has stayed level in the United States, while it has declined in Europe. Today, the typical American works many more hours than the typical resident of these two western European countries.

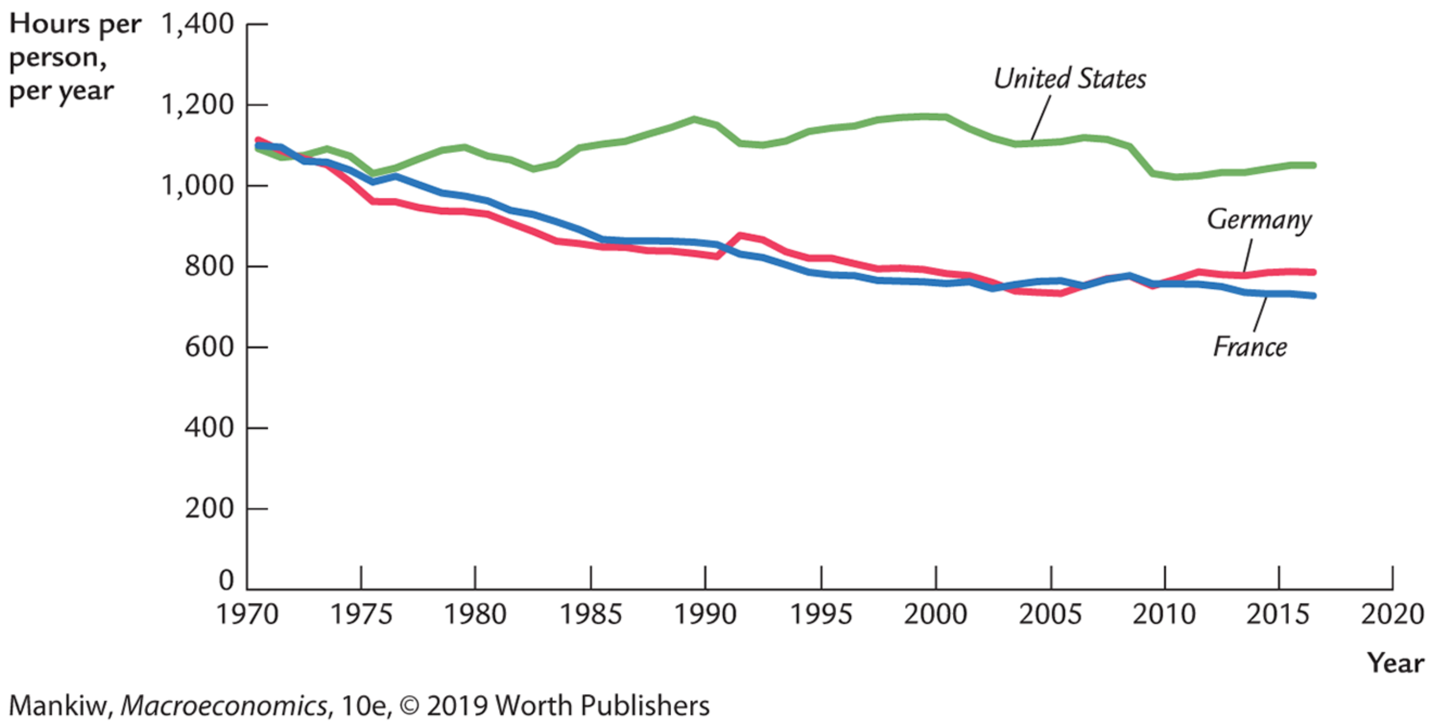


FIGURE 7-7 Annual Hours Worked per Person Over time, many Europeans have substantially reduced the number of hours they work, while typical Americans have not.

Data from: OECD and Bureau of Labor Statistics. Calculated as the average annual hours actually worked per employed person multiplied by the employment rate.

The difference in hours worked reflects two facts. First, the average employed person in the United States works more hours per year than the average employed person in Europe. Europeans typically enjoy shorter workweeks and more holidays. Second, more potential workers are employed in the United States. That is, the employment-to-population ratio is higher in the United States than it is in Europe. Higher unemployment is one reason for the lower employment-to-population ratio in Europe. Another reason is earlier retirement in Europe and thus lower labor-force participation among older workers.

What is the underlying cause of these differences in work patterns? Economists have proposed several hypotheses.

Edward Prescott, the 2004 winner of the Nobel Prize in economics, has concluded that “virtually all of the large differences between U.S. labor supply and those of Germany and France are due to differences in tax systems.” This hypothesis is consistent with two facts: (1) Europeans face higher tax rates than Americans, and (2) European tax rates have risen significantly over the past several decades. Some economists take these facts as evidence for the impact of taxes on work effort. Others are skeptical. They argue that explaining the difference in hours worked by tax rates alone requires an implausibly large elasticity of labor supply.

A related hypothesis is that the difference in observed work effort may be attributable to the underground economy. When tax rates are high, people have a greater incentive to work “off the books” to evade taxes. Data on the underground economy are hard to come by, but economists who study the subject believe the underground economy is larger in Europe than it is in the United States. This fact suggests that the difference in actual hours worked, including work in the underground economy, may be smaller than the difference in measured hours worked.

Another hypothesis stresses the role of unions. As we have seen, collective bargaining is more important in European than in U.S. labor markets. Unions often push for shorter workweeks in contract negotiations, and they lobby the government for various labor-market regulations, such as official holidays. Economists Alberto Alesina, Edward Glaeser, and Bruce Sacerdote conclude that "mandated holidays can explain 80 percent of the difference in weeks worked between the U.S. and Europe and 30 percent of the difference in total labor supply between the two regions." They suggest that Prescott overstates the role of taxes because, looking across countries, tax rates and unionization rates are positively correlated; as a result, the effects of taxes and the effects of unionization are hard to disentangle.

A final hypothesis emphasizes the possibility of different preferences. As technological progress and economic growth have made countries richer, people around the world must decide whether to take the greater prosperity in the form of increased consumption of goods and services or increased leisure. According to economist Olivier Blanchard, "the main difference [between the continents] is that Europe has used some of the increase in productivity to increase leisure rather than income, while the U.S. has done the opposite." Blanchard believes that Europeans simply have more taste for leisure than do Americans. (As a French economist working in the United States, he may have special insight into this phenomenon.) If Blanchard is right, this raises the harder question of why tastes vary by geography.

Economists debate the merits of these alternative hypotheses. In the end, there may be some truth to all of them.[11]

7-6 Conclusion

Unemployment represents wasted resources. Unemployed workers have the potential to contribute to national income but are not doing so. Those searching for jobs to suit their skills are happy when the search is over, and those waiting for jobs that pay above-equilibrium wages are happy when positions open up.

Unfortunately, neither frictional unemployment nor structural unemployment can be easily reduced. The government cannot make job search instantaneous, and it cannot easily bring wages closer to equilibrium levels. Zero unemployment is not a plausible goal for free-market economies.

Yet public policy is not powerless in the fight to reduce unemployment. Job-training programs, the unemployment-insurance system, the minimum wage, and the laws governing collective bargaining are often topics of political debate. The policies we choose are likely to have important effects on the economy's natural rate of unemployment.

CHAPTER 8

Economic Growth I: Capital Accumulation and Population Growth

Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

> The question of growth is nothing new but a new disguise for an age-old issue, one which has always intrigued and preoccupied economics: the present versus the future.
>
> —James Tobin

If you have ever spoken with your grandparents about what their lives were like when they were young, most likely you learned an important lesson about economic growth: material standards of living have improved substantially over time for most families in most countries. This advance comes from rising incomes, which have allowed people to consume greater quantities of goods and services.

To measure economic growth, economists use data on gross domestic product, which measures the total income of everyone in a nation's economy. The real GDP of the United States today is more than seven times its 1950 level, and real GDP per person is more than three times its 1950 level. In any given year, we also observe large differences in the standard of living among countries. Table 8-1 shows the 2016 income per person in the world's 14 most populous countries. The United States tops the list, with an income of $57,467 per person. Ethiopia has an income per person of only $1,735—about 3 percent of the figure for the United States.

TABLE 8-1 International Differences in the Standard of Living

Country	Income per Person (2016)
United States	$57,467
Japan	41,470
Russia	23,163
Mexico	17,862

China	15,535
Brazil	15,128
Indonesia	11,612
Egypt	11,132
Philippines	7,806
India	6,572
Nigeria	5,867
Pakistan	5,249
Bangladesh	3,581
Ethiopia	1,735
Data from: The World Bank. Data are PPP-adjusted—that is, the income figures account for differences in the cost of living among countries.	

Our goal in this part of the book is to understand what causes these differences in income over time and across countries. In Chapter 3 we identified the factors of production—capital and labor—and the production technology as the sources of the economy's output and, thus, of its total income. Differences in income across time and across nations must then come from differences in capital, labor, and technology.

Our main task in this chapter and the next is to develop a theory of economic growth called the **Solow growth model**. Our analysis in Chapter 3 enabled us to describe how the economy produces and uses its output at a point in time. The analysis was static—a snapshot of the economy. To explain why national income grows, and why some economies grow faster than others, we must broaden our analysis so that it describes changes in the economy over time. By developing such a model, we make our analysis dynamic—more like a movie than a photograph. The Solow model shows how saving, population growth, and technological progress affect the level of an economy's output and its growth over time. In this chapter we analyze the roles of saving and population growth. In the next chapter we introduce technological progress.[1]

8-1 The Accumulation of Capital

The Solow model is designed to show how growth in the capital stock, growth in the labor force, and advances in technology interact in an economy as well as how they affect a nation's total output of goods and services. We will build this model in a series of steps. Our first step is to examine how the supply and demand for goods determine the accumulation of capital. In this first step, we assume that the labor force and technology are fixed. We then relax these assumptions by introducing changes in the labor force later in this chapter and changes in technology in the next.

The Supply and Demand for Goods

The supply and demand for goods played a central role in our static model of the closed economy in Chapter 3. The same is true for the Solow model. By considering the supply and demand for goods, we can see what determines how much output is produced at any given time and how this output is allocated among alternative uses.

The Supply of Goods and the Production Function

The supply of goods in the Solow model is based on the production function, which states that output depends on the capital stock and the labor force:

$$Y = F(K, L).$$

The Solow model assumes that the production function has constant returns to scale. This assumption is often considered realistic, and, as we will see shortly, it simplifies the analysis. Recall that a production function has constant returns to scale if

$$zY = F(zK, zL)$$

for any positive number z. That is, if both capital and labor are multiplied by z, the amount of output is also multiplied by z.

Production functions with constant returns to scale allow us to analyze all quantities in the economy relative to the size of the labor force. To see that this is true, set $z = 1/L$ in the preceding equation to obtain

$$Y/L = F(K/L, 1).$$

This equation shows that the amount of output per worker Y/L is a function of the amount of capital per worker K/L. (The number 1 is constant and thus can be ignored.) The assumption of constant returns to scale implies that the size of the economy—as measured by the number of workers—does not affect the relationship between output per worker and capital per worker.

Because the size of the economy does not matter, it will prove convenient to denote all quantities in per-worker terms. We designate quantities per worker with lowercase letters, so $y = Y/L$ is output per worker, and $k = K/L$ is capital per worker. We can then write the production function as

$$y = f(k),$$

where we define $f(k) = F(k, 1)$. Figure 8-1 illustrates this production function.

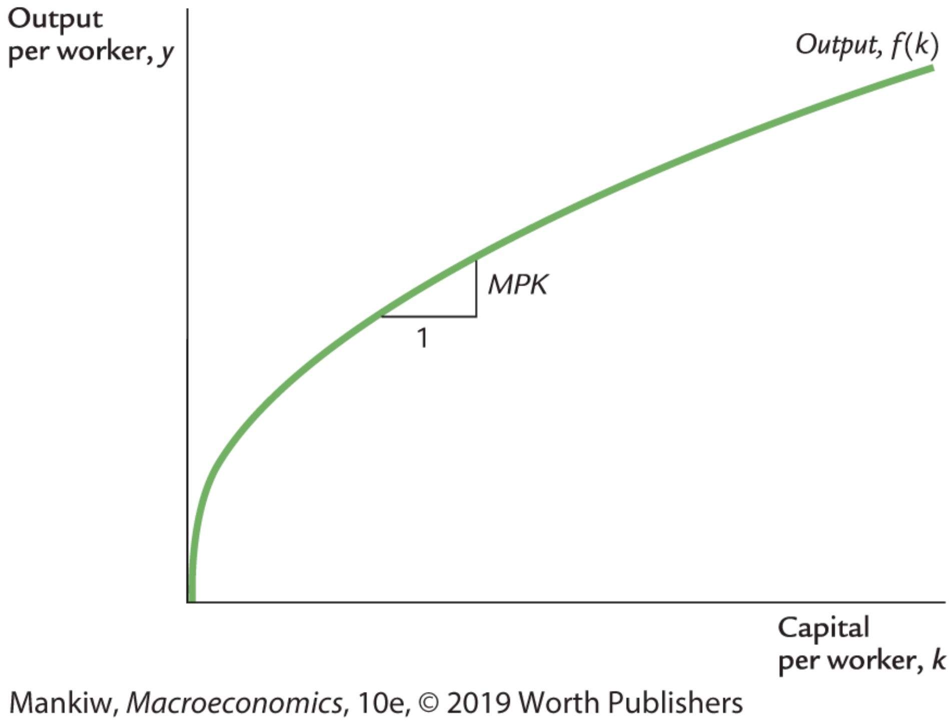


FIGURE 8-1 The Production Function The production function shows how the amount of capital per worker *k* determines the amount of output per worker $y = f(k)$. The slope of the production function is the marginal product of capital: if *k* increases by 1 unit, *y* increases by *MPK* units. The production function becomes flatter as *k* increases, indicating diminishing marginal product of capital.

The slope of this production function shows how much extra output a worker produces when given an extra unit of capital. This amount is the marginal product of capital *MPK*. Mathematically, we write

$$MPK = f(k+1) - f(k).$$

Note that in Figure 8-1, as the amount of capital increases, the production function becomes flatter, indicating that the production function exhibits diminishing marginal product of capital. When *k* is low, the average worker has only a little capital to work with, so an extra unit of capital is very useful and produces a lot of additional output. When *k* is high, the average worker has a lot of capital already, so an extra unit increases production only slightly.

The Demand for Goods and the Consumption Function

The demand for goods in the Solow model comes from consumption and investment. In other words, output per worker *y* is divided between consumption per worker *c* and investment per worker *i*:

$$y = c + i.$$

This equation is the per-worker version of the economy's national income accounts identity. It omits government purchases (which for present purposes we can ignore) and net exports (because we are assuming a closed economy).

The Solow model assumes that each year people save a fraction *s* of their income and consume a fraction $(1-s)$. We can express this idea with the following consumption function:

$$c = (1-s)y,$$

where *s*, the saving rate, is a number between zero and one. Keep in mind that government policies can potentially influence a nation's saving rate, so one of our goals is to find what saving rate is desirable. For now, however, we just take the saving rate *s* as given.

To see what this consumption function implies for investment, substitute $(1-s)y$ for *c* in the national income accounts identity:

$$y = (1 - s)y + i.$$

Rearrange the terms to obtain

$$i = sy.$$

This equation shows that investment equals saving, as we first saw in Chapter 3. Thus, the rate of saving *s* is also the fraction of output devoted to investment.

We have now introduced the two main ingredients of the Solow model—the production function and the consumption function—which describe the economy at any moment in time. For any given capital stock *k*, the production function $y = f(k)$ determines how much output the economy produces, and the saving rate *s* determines the allocation of that output between consumption and investment.

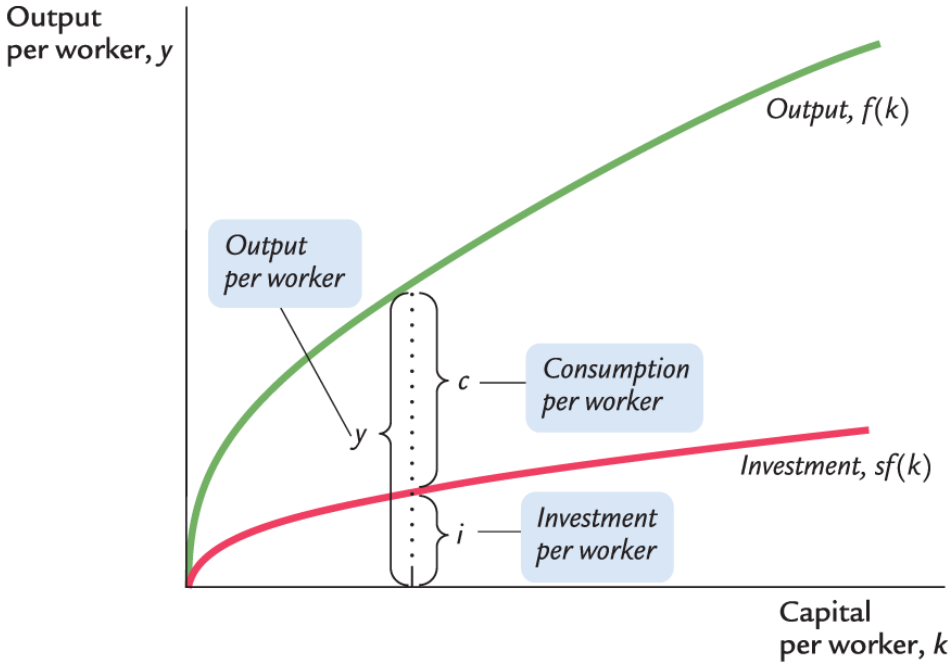
Growth in the Capital Stock and the Steady State

At any moment, the capital stock is a key determinant of the economy's output, but the capital stock can change over time, and those changes can lead to economic growth. Two forces influence the capital stock: investment and depreciation. *Investment* is expenditure on new plant and equipment, and it causes the capital stock to rise. *Depreciation* is the wearing out of old capital due to aging and use, and it causes the capital stock to fall. Let's consider each of these forces in turn.

As we have already noted, investment per worker *i* equals *sy*. By substituting the production function for *y*, we can express investment per worker as a function of the capital stock per worker:

$$i = sf(k).$$

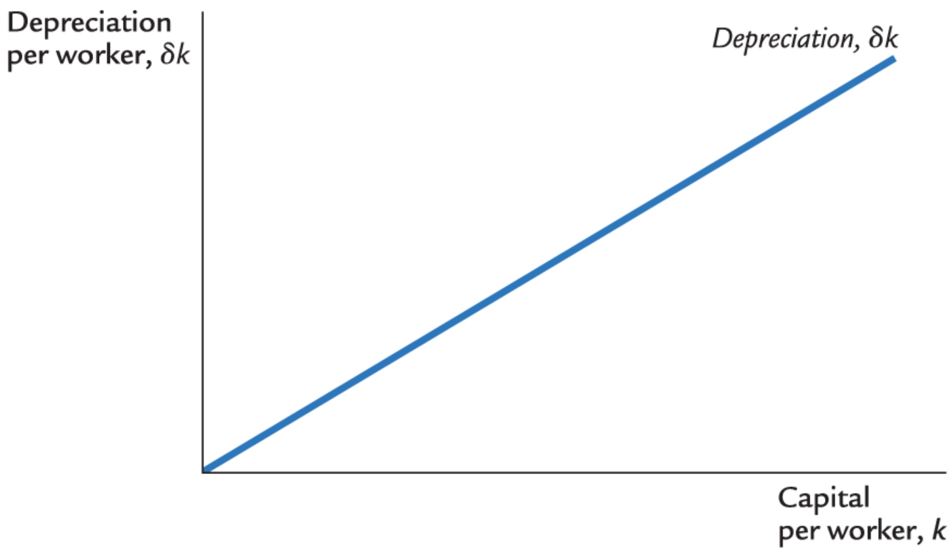
This equation relates the existing stock of capital *k* to the accumulation of new capital *i*. Figure 8-2 shows this relationship. This figure illustrates how, for any value of *k*, the amount of output is determined by the production function $f(k)$, and the allocation of that output between consumption and investment is determined by the saving rate *s*.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 8-2 Output, Consumption, and Investment The saving rate s determines the allocation of output between consumption and investment. For any level of capital k, output is $f(k)$, investment is $sf(k)$, and consumption is $f(k) - sf(k)$.

To incorporate depreciation into the model, we assume that a certain fraction δ of the capital stock wears out each year. Here δ (the lowercase Greek letter delta) is called the *depreciation rate*. For example, if capital lasts an average of 25 years, the depreciation rate is 4 percent per year $(\delta = 0.04)$. The amount of capital that depreciates each year is δk. Figure 8-3 shows how the amount of depreciation depends on the capital stock.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 8-3 Depreciation A constant fraction δ of the capital stock wears out every year. Depreciation is therefore proportional to the capital stock.

We can express the impact of investment and depreciation on the capital stock with this equation:

$$\begin{array}{lcccc} \text{Change in Capital Stock} & = & \text{Investment} & - & \text{Depreciation} \\ \Delta k & = & i & - & \delta k, \end{array}$$

where Δk is the change in the capital stock between one year and the next. Because investment *i* equals *sf*(*k*), we can write this as

$$\Delta k = sf(k) - \delta k.$$

Figure 8-4 graphs the terms of this equation—investment and depreciation—for different levels of the capital stock *k*. The higher the capital stock, the greater the amounts of output and investment. Yet the higher the capital stock, the greater also the amount of depreciation.

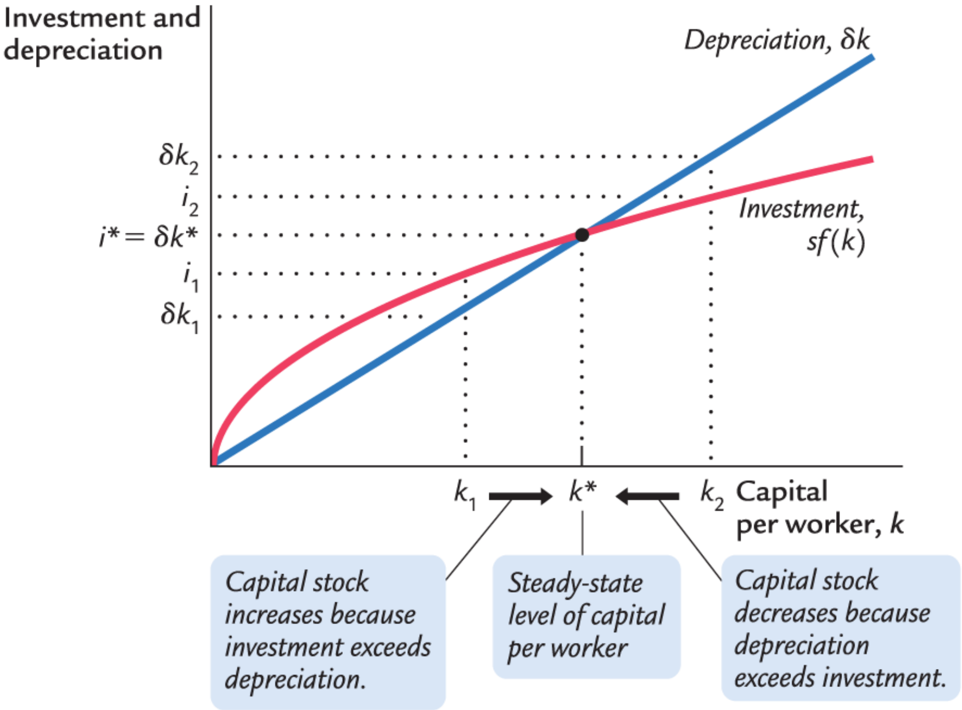


FIGURE 8-4 Investment, Depreciation, and the Steady State The steady-state level of capital k^* is the level at which investment equals depreciation, indicating that the amount of capital will not change over time. Below k^*, investment exceeds depreciation, so the capital stock grows. Above k^*, investment is less than depreciation, so the capital stock shrinks.

As Figure 8-4 shows, there is a single capital stock k^* at which the amount of investment equals the amount of depreciation. If the economy finds itself at this level of the capital stock, the capital stock will not change because the two forces acting on it—investment and depreciation—just balance. That is, at k^*, $\Delta k = 0$, so the capital stock *k* and output *f*(*k*) are steady over time (rather than growing or shrinking). We therefore call k^* the *steady-state level of capital*.

The **steady state** is significant for two reasons. As we have just seen, an economy at the steady state will stay there. In addition, and just as important, an economy *not* at the steady state will go there. That is, regardless of the level of capital with which the economy begins, it ends up with the steady-state level of capital. In this sense, *the steady state represents the long-run equilibrium of the economy.*

To see why an economy always ends up at the steady state, suppose that the economy starts with less than the steady-state level of capital, such as level k_1 in Figure 8-4. In this case, investment exceeds depreciation. Over time, the capital stock will rise and will continue to rise—along with output *f*(*k*)—until it approaches the steady state k^*.

Similarly, suppose that the economy starts with more than the steady-state level of capital, such as level k_2. In this case, investment is less than depreciation: capital is wearing out faster than it is being replaced. The capital stock will fall, again approaching the steady-state level. Once the capital stock reaches the steady state, investment equals depreciation, and there is no pressure for the capital stock to either increase or decrease.

Approaching the Steady State: A Numerical Example

Let's use a numerical example to see how the Solow model works and how the economy approaches the steady state. For this example, we assume that the production function is

$$Y = K^{1/2} L^{1/2}.$$

From Chapter 3, you will recognize this as the Cobb–Douglas production function with the capital-share parameter α equal to 1/2. To derive the per-worker production function *f*(*k*), divide both sides of the production function by the labor force *L*:

$$\frac{Y}{L} = \frac{K^{1/2} L^{1/2}}{L}.$$

Rearrange to obtain

$$\frac{Y}{L} = \left(\frac{K}{L}\right)^{1/2}.$$

Because $y = Y/L$ and $k = K/L$, this equation becomes

$$y = k^{1/2},$$

which can also be written as

$$y = \sqrt{k}.$$

This form of the production function states that output per worker equals the square root of the amount of capital per worker.

To complete the example, let's assume that 30 percent of output is saved $(s = 0.3)$, 10 percent of the capital stock depreciates every year $(\delta = 0.1)$, and the economy starts off with 4 units of capital per worker $(k = 4)$. Given these numbers, we can now examine what happens to this economy over time.

We begin by looking at the production and allocation of output in the first year, when the economy has 4 units of capital per worker. Here are the steps we follow.

- According to the production function $y = \sqrt{k}$, the 4 units of capital per worker (*k*) produce 2 units of output per worker (*y*).
- Because 30 percent of output is saved and invested and 70 percent is consumed, $i = 0.6$ and $c = 1.4$.
- Because 10 percent of the capital stock depreciates, $\delta k = 0.4$.
- With investment of 0.6 and depreciation of 0.4, the change in the capital stock is $\Delta k = 0.2$.

Thus, the economy begins its second year with 4.2 units of capital per worker.

We can do the same calculations for each subsequent year. Table 8-2 shows how the economy progresses. Every year, because investment exceeds depreciation, new capital is added, and output grows. Over many years, the economy approaches a steady state with 9 units of capital per worker. In this steady state, investment of 0.9 exactly offsets depreciation of 0.9, so the capital stock and output are no longer growing.

TABLE 8-2 Approaching the Steady State: A Numerical Example

Assumptions: $y=k$; $s=0.3$; $\delta=0.1$; initial $k=4.0$

Year	k	y	c	i	δk	Δk
1	4.000	2.000	1.400	0.600	0.400	0.200
2	4.200	2.049	1.435	0.615	0.420	0.195
3	4.395	2.096	1.467	0.629	0.440	0.189
4	4.584	2.141	1.499	0.642	0.458	0.184
5	4.768	2.184	1.529	0.655	0.477	0.178
.						
.						
.						
10	5.602	2.367	1.657	0.710	0.560	0.150
.						
.						
.						
25	7.321	2.706	1.894	0.812	0.732	0.080
.						
.						
.						
100	8.962	2.994	2.096	0.898	0.896	0.002
.						
.						
.						
∞	9.000	3.000	2.100	0.900	0.900	0.000

Following the progress of the economy for many years is one way to find the steady-state capital stock, but there is another way that requires fewer calculations. Recall that

Δk=sf(k)−δk.

$$\Delta k = sf(k) - \delta k.$$

This equation shows how *k* evolves over time. Because the steady state is (by definition) the value of *k* at which $\Delta k = 0$, we know that

$$0 = sf(k^*) - \delta k^*,$$

or, equivalently,

$$\frac{k^*}{f(k^*)} = \frac{s}{\delta}.$$

This equation provides a way of finding the steady-state level of capital per worker k^*. Substituting in the numbers and production function from our example, we obtain

$$\frac{k^*}{\sqrt{k^*}} = \frac{0.3}{0.1}.$$

Now square both sides of this equation to find

$$k^* = 9.$$

The steady-state capital stock is 9 units per worker. This result confirms the calculation of the steady state in Table 8-2.

CASE STUDY

The Miracle of Japanese and German Growth

Japan and Germany are two success stories of economic growth. Although today they are economic superpowers, in 1946 both countries faced economies in shambles. World War II had destroyed much of their capital stocks. In both nations, output per person in 1946 was about half of what it had been before the war. In the following decades, however, these two countries experienced some of the most rapid growth rates on record. Between 1946 and 1972, output per person grew at 8.0 percent per year in Japan and 6.5 percent per year in Germany, compared to only 2.1 percent per year in the United States. Several other European economies damaged by the war also enjoyed rapid growth during this postwar period: for example, output per worker grew at 4.6 percent per year in France and 5.5 percent per year in Italy. But Japan and Germany are the two nations that experienced both the greatest devastation during the war and the most rapid growth afterward.

Are these postwar experiences surprising from the standpoint of the Solow model? Consider an economy in steady state. Now suppose that a war destroys some of the capital stock. (That is, suppose the capital stock drops from k^* to k_1 in Figure 8-4.) Not surprisingly, the level of output falls immediately. But if the saving rate—the fraction of output devoted to saving and investment—is unchanged, the economy will then experience a period of high growth. Output grows because, at the lower capital stock, more capital is added by investment than is removed by depreciation. This high growth continues until the economy approaches its former steady state. Hence, although destroying part of the capital stock immediately reduces output, it is followed by higher-than-normal growth. The "miracle" of rapid growth in Japan and Germany, as it is often described in the business press, is what the Solow model predicts for countries in which war has greatly reduced the capital stock.

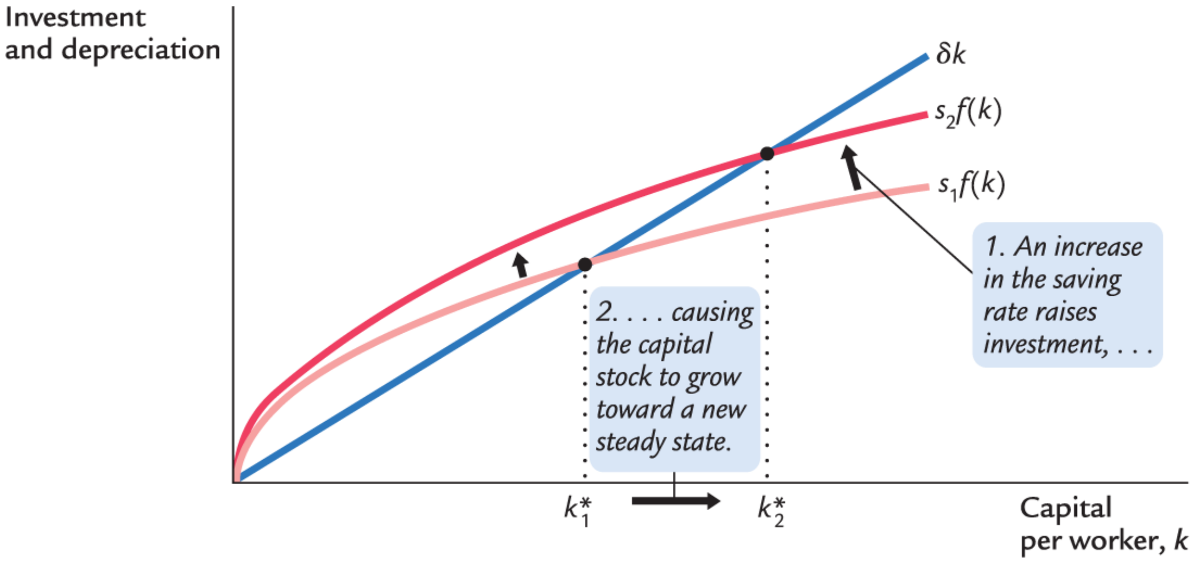
After their postwar growth miracles, both Japan and Germany settled down to moderate rates of growth, more like that of the United States. From 1972 to 2000, output per person grew at 2.4 percent per year in Japan and 1.8 percent per year in Germany, compared to 2.1 percent per year in the United States. This phenomenon is also what the Solow model predicts. As an economy gets closer to its steady state, it no longer experiences the higher-than-normal growth that arises from the transition back to the steady state.

Lest one take the wrong lesson from this historical episode, note that wartime destruction should not be seen as desirable. The fast growth in Japan and Germany during the postwar period merely caught them up to where they otherwise would have been. Moreover, unlike Japan and Germany, many war-torn nations are left with a legacy of civil strife and political instability, hampering their subsequent growth. ■

How Saving Affects Growth

The explanation of Japanese and German growth after World War II is not quite as simple as suggested in the preceding Case Study. Another relevant fact is that both Japan and Germany save and invest a higher fraction of their output than the United States. To understand more fully the international differences in economic performance, we must consider the effects of different saving rates.

Consider what happens to an economy when its saving rate increases. Figure 8-5 shows such a change. The economy is assumed to begin in a steady state with saving rate s_1 and capital stock k_1^*. When the saving rate increases from s_1 to s_2, the $sf(k)$ curve shifts upward. At the initial saving rate s_1 and the initial capital stock k_1^*, the amount of investment just offsets the amount of depreciation. Immediately after the saving rate rises, investment is higher, but the capital stock and depreciation are unchanged. Therefore, investment exceeds depreciation. The capital stock gradually rises until the economy reaches the new steady state k_2^*, which has a higher capital stock and a higher level of output than the old steady state.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 8-5 An Increase in the Saving Rate An increase in the saving rate *s* implies that the amount of investment for any given capital stock is higher. It therefore shifts the saving function upward. At the initial steady state, k1*, k_1^*, investment now exceeds depreciation. The capital stock rises until the economy reaches a new steady state k2* k_2^* with more capital and output.

The Solow model shows that the saving rate is a key determinant of the steady-state capital stock. *If the saving rate is high, the economy will have a large capital stock and a high level of output in the steady state. If the saving rate is low, the economy will have a small capital stock and a low level of output in the steady state*. This conclusion sheds light on many discussions of fiscal policy. As we saw in Chapter 3, a government budget deficit can reduce national saving and crowd out investment. Now we can see that the long-run consequences of a reduced saving rate are a lower capital stock and lower national income. This is why many economists are critical of persistent budget deficits.

What does the Solow model say about the relationship between saving and economic growth? Higher saving leads to faster growth in the Solow model—but only temporarily. An increase in the saving rate raises growth only until the economy reaches the new steady state. If the economy maintains a high saving rate, it will maintain a large capital stock and a high level of output, but it will not maintain a high growth rate forever. Policies that alter the steady-state growth rate of income per person are said to have a *growth effect*; we will see examples in the next chapter. By contrast, a higher saving rate is said to have a *level effect* because only the level of income per person—not its growth rate—is influenced by the saving rate in the steady state.

Now that we understand how saving and growth interact, we can more fully explain the impressive economic performance of Germany and Japan after World War II. Not only were their initial capital stocks low because of the war but their steady-state capital stocks were also high because of their high saving rates. Both facts help explain the rapid growth of these two countries in the 1950s and 1960s.

8-2 The Golden Rule Level of Capital

So far, we have used the Solow model to examine how an economy's rate of saving and investment determines its steady-state levels of capital and income. This analysis might lead you to think that higher saving is always a good thing because it leads to greater income. Yet suppose a nation had a saving rate of 100 percent. That would lead to the largest possible capital stock and the largest possible income. But if all of this income is saved and none is ever consumed, what good is it?

This section uses the Solow model to discuss the optimal amount of capital accumulation from the standpoint of economic well-being. In the next chapter, we discuss how government policies influence a nation's saving rate. But first, in this section, we present the theory behind these policy decisions.

Comparing Steady States

To keep our analysis simple, let's assume that a policymaker can set the economy's saving rate at any level. By setting the saving rate, the policymaker determines the economy's steady state. What steady state should the policymaker choose?

The policymaker's goal is to maximize the well-being of the individuals who make up the society. Individuals themselves do not care about the amount of capital in the economy or even the amount of output. They care about the amount of goods and services they can consume. Thus, a benevolent policymaker would want to choose the steady state with the highest level of consumption. The steady-state value of *k* that maximizes consumption is called the **Golden Rule level of capital** and is denoted k^*_{gold}.[2]

How can we tell whether an economy is at the Golden Rule level? To answer this question, we must first determine steady-state consumption per worker. Then we can see which steady state provides the most consumption.

To find steady-state consumption per worker, we begin with the national income accounts identity

$$y = c + i$$

and rearrange it as

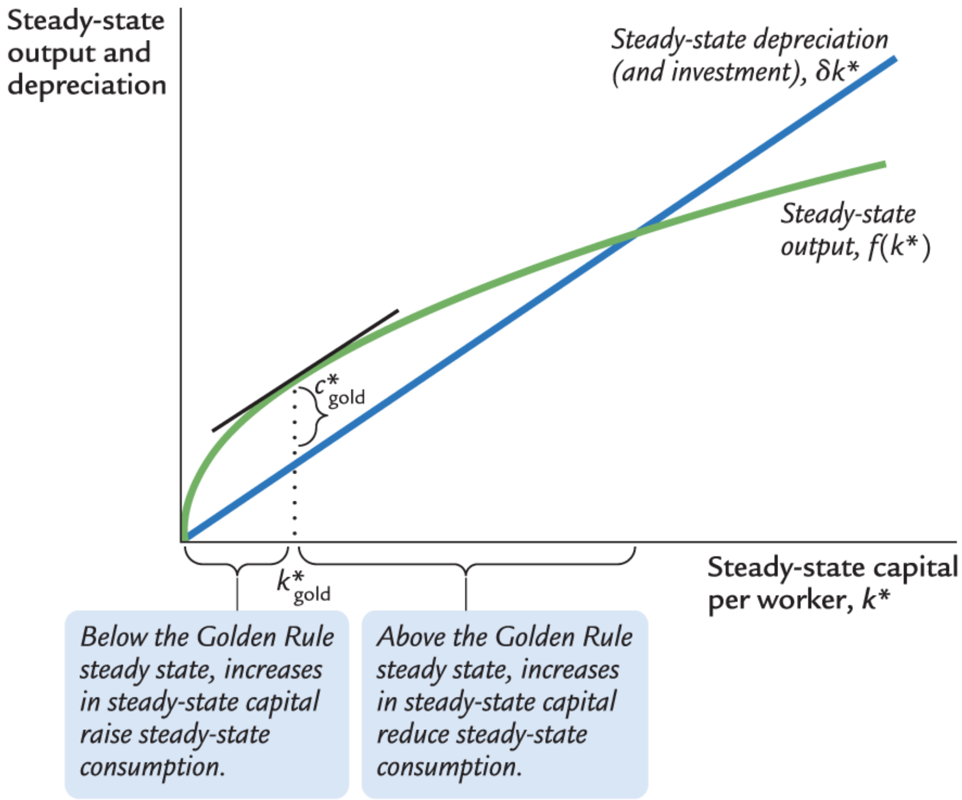
$$c = y - i.$$

Consumption is output minus investment. Because we want to find steady-state consumption, we substitute steady-state values for output and investment. Steady-state output per worker is $f(k^*)$, where k^* is the steady-state capital stock per worker. Furthermore, because the capital stock is not changing in the steady state, investment equals depreciation δk^*. Substituting $f(k^*)$ for *y* and δk^* for *i*, we can write steady-state consumption per worker as

$$c^* = f(k^*) - \delta k^*.$$

According to this equation, steady-state consumption is what's left of steady-state output after paying for steady-state depreciation. This equation shows that an increase in steady-state capital has two opposing effects on steady-state consumption. On the one hand, more capital means more output. On the other hand, more capital also means that more output must be used to replace capital that is wearing out.

Figure 8-6 graphs steady-state output and steady-state depreciation as a function of the steady-state capital stock. Steady-state consumption is the gap between output and depreciation. This figure shows that there is one level of the capital stock—the Golden Rule level k^*_{gold}—that maximizes consumption.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 8-6 Steady-State Consumption The economy's output is used for consumption or investment. In the steady state, investment equals depreciation. Therefore, steady-state consumption is the difference between output $f(k^*)$ and depreciation δk^*. Steady-state consumption is maximized at the Golden Rule steady state. The Golden Rule capital stock is denoted k^*_{gold}, and the Golden Rule level of consumption is denoted c^*_{gold}.

When comparing steady states, we must keep in mind that higher levels of capital affect both output and depreciation. If the capital stock is below the Golden Rule level, an increase in the capital stock raises output more than depreciation, so consumption rises. In this case, the production function is steeper than the δk^* line, so the gap between these two curves—which equals consumption—grows as k^* rises. By contrast, if the capital stock is above the Golden Rule level, an increase in the capital stock reduces consumption because the increase in output is smaller than the increase in depreciation. In this case, the production function is flatter than the δk^* line, so the gap between the curves—consumption—shrinks as k^* rises. At the Golden Rule level of capital, the production function and the δk^* line have the same slope, and consumption is at its greatest level.

We can now derive a simple condition that characterizes the Golden Rule level of capital. Recall that the slope of the production function is the marginal product of capital *MPK*. The slope of the δk^* line is δ. Because these two slopes are equal at k^*_{gold}, the Golden Rule is described by the equation

$$MPK = \delta.$$

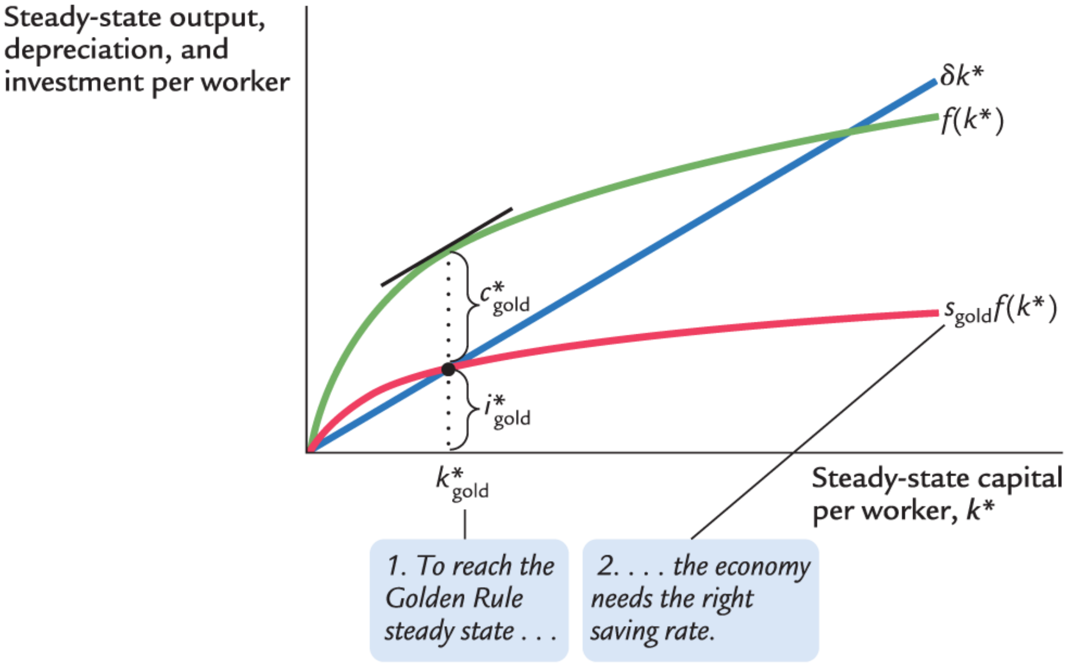
At the Golden Rule level of capital, the marginal product of capital equals the depreciation rate.

To make the point somewhat differently, suppose that the economy starts at some steady-state capital stock k^* and that the policymaker is considering increasing the capital stock to $k^* + 1$. The amount of extra output from this increase in capital would be $f(k^* + 1) - f(k^*)$, the marginal product of capital, *MPK*. The amount of extra depreciation from having 1 more unit of capital is the depreciation rate δ. Thus, the net effect of this extra unit of capital on consumption is $MPK - \delta$. If $MPK - \delta > 0$, increases in capital increase consumption, so k^* must be below the Golden Rule level. If $MPK - \delta < 0$, increases in capital decrease consumption, so k^* must be above the Golden Rule level. Therefore, the following condition describes the Golden Rule:

$$MPK - \delta = 0.$$

At the Golden Rule level of capital, the marginal product of capital net of depreciation $(MPK - \delta)$ equals zero. As we will see, a policymaker can use this condition to find the Golden Rule capital stock for an economy.[3]

Keep in mind that the economy does not automatically gravitate toward the Golden Rule steady state. If we want any particular steady-state capital stock, such as the Golden Rule, we need a particular saving rate to support it. Figure 8-7 shows the steady state if the saving rate is set to produce the Golden Rule level of capital. If the saving rate is higher than the one used in this figure, the steady-state capital stock will be too high. If the saving rate is lower, the steady-state capital stock will be too low. In either case, steady-state consumption will be lower than it is at the Golden Rule steady state.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 8-7 The Saving Rate and the Golden Rule There is only one saving rate that produces the Golden Rule level of capital k^*_{gold}. Any change in the saving rate would shift the *sf*(*k*) curve and would move the economy to a steady state with a lower level of consumption.

Finding the Golden Rule Steady State: A Numerical Example

Consider a policymaker choosing a steady state in the following economy. The production function is the same as in our earlier example:

$$y = \sqrt{k}.$$

Output per worker is the square root of capital per worker. Depreciation δ is again 10 percent of capital. This time, the policymaker picks the saving rate *s* and thus the economy's steady state.

To see the outcomes available to the policymaker, recall that the following equation holds in the steady state:

k*f(k*)=sδ. $$\frac{k^*}{f(k^*)} = \frac{s}{\delta}.$$

In this economy, this equation becomes

k*k*=s0.1. $$\frac{k^*}{\sqrt{k^*}} = \frac{s}{0.1}.$$

Squaring both sides of this equation yields the steady-state capital stock:

k*=100s2. $$k^* = 100s^2.$$

Using this result, we can compute the steady-state capital stock for any saving rate.

Table 8-3 presents calculations showing the steady states that result from various saving rates. We see that higher saving leads to a higher capital stock, increasing both output and depreciation. Steady-state consumption, the difference between output and depreciation, first rises with higher saving rates and then declines. Consumption is highest when the saving rate is 0.5. Hence, a saving rate of 0.5 produces the Golden Rule steady state.

TABLE 8-3 Finding the Golden Rule Steady State: A Numerical Example

Assumptions: y=k;δ=0.1

s	k*	y*	δk*	c*	MPK	MPK−δ
0.0	0.0	0.0	0.0	0.0	∞	∞
0.1	1.0	1.0	0.1	0.9	0.500	0.400
0.2	4.0	2.0	0.4	1.6	0.250	0.150
0.3	9.0	3.0	0.9	2.1	0.167	0.067
0.4	16.0	4.0	1.6	2.4	0.125	0.025
0.5	**25.0**	**5.0**	**2.5**	**2.5**	**0.100**	**0.000**
0.6	36.0	6.0	3.6	2.4	0.083	−0.017
0.7	49.0	7.0	4.9	2.1	0.071	−0.029

0.8	64.0	8.0	6.4	1.6	0.062	−0.038
0.9	81.0	9.0	8.1	0.9	0.056	−0.044
1.0	100.0	10.0	10.0	0.0	0.050	−0.050

Recall that another way to identify the Golden Rule steady state is to find the capital stock at which the net marginal product of capital $(MPK - \delta)$ equals zero. For this production function, the marginal product is[4]

$$MPK = \frac{1}{2\sqrt{k}}.$$

Using this formula, the last two columns of Table 8-3 present the values of *MPK* and $MPK - \delta$ in the different steady states. Note that the net marginal product of capital is exactly zero when the saving rate is at its Golden Rule value of 0.5. Because of diminishing marginal product, the net marginal product of capital is greater than zero whenever the economy saves less than this amount, and it is less than zero whenever the economy saves more.

This numerical example confirms that the two ways of finding the Golden Rule steady state—looking at steady-state consumption or looking at the marginal product of capital—give the same answer. If we want to know whether an actual economy is currently at, above, or below its Golden Rule capital stock, the second method is more convenient because it is straightforward to estimate the marginal product of capital. By contrast, evaluating an economy with the first method requires estimates of steady-state consumption at many different saving rates; such information is harder to obtain. Thus, when we apply this kind of analysis to the U.S. economy in the next chapter, we will evaluate U.S. saving by examining the marginal product of capital. Before doing so, however, we need to proceed further in our development of the Solow model.

The Transition to the Golden Rule Steady State

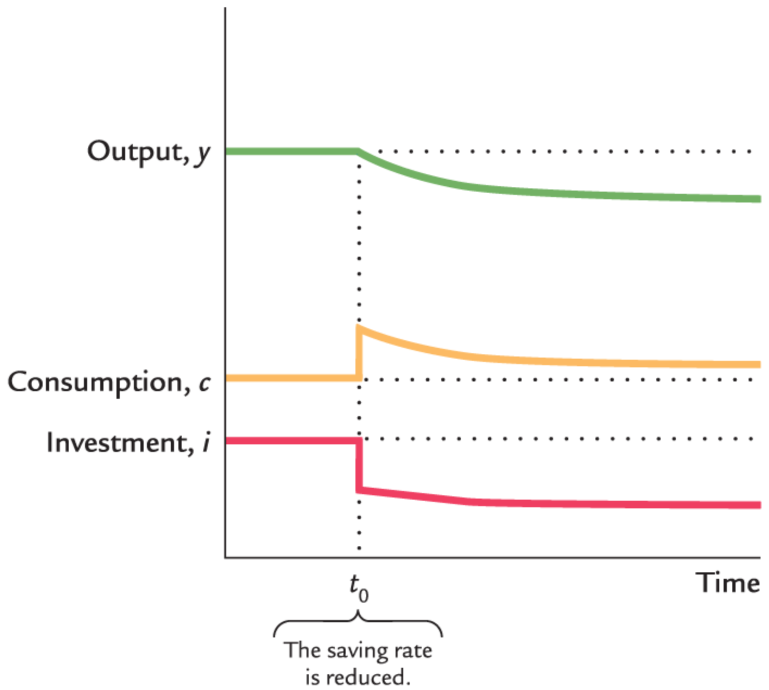
Let’s now make our policymaker’s problem more realistic. So far, we have been assuming that the policymaker can simply choose the economy’s steady state and jump there immediately. In this case, the policymaker would choose the steady state with the highest consumption: the Golden Rule steady state. But now suppose that the economy has reached a steady state other than the Golden Rule. What happens to consumption, investment, and capital when the economy makes the transition between steady states? Might the impact of the transition deter the policymaker from trying to achieve the Golden Rule?

We must consider two cases: the economy might begin with more capital than in the Golden Rule steady state, or it might begin with less. It turns out that the two cases offer very different problems for policymakers. (As we will see in the next chapter, the second case—too little capital—describes most actual economies, including that of the United States.)

Starting with Too Much Capital

We first consider the case in which the economy begins at a steady state with more capital than it would have in the Golden Rule steady state. In this case, the policymaker should pursue policies aimed at reducing the rate of saving in order to reduce the capital stock. Suppose that these policies succeed and that at some point—call it time t_0—the saving rate falls to the level that will lead to the Golden Rule steady state.

Figure 8-8 shows what happens to output, consumption, and investment when the saving rate falls. The reduction in the saving rate causes an immediate increase in consumption and a decrease in investment. Because investment and depreciation were equal in the initial steady state, investment will now be less than depreciation, which means the economy is no longer in a steady state. Gradually, the capital stock falls, leading to reductions in output, consumption, and investment. These variables continue to fall until the economy reaches the new steady state. Because we are assuming that the new steady state is the Golden Rule steady state, consumption must be higher than it was before the change in the saving rate, even though output and investment are lower.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 8-8 Reducing Saving When Starting with More Capital Than in the Golden Rule Steady State This figure shows what happens over time to output, consumption, and investment when the economy begins with more capital than the Golden Rule level and the saving rate is reduced. The reduction in the saving rate (at time t0t_0) causes an immediate increase in consumption and an equal decrease in investment. Over time, as the capital stock falls, output, consumption, and investment fall together. Because the economy began with too much capital, the new steady state has a higher level of consumption than the initial steady state.

Note that, compared to consumption in the old steady state, consumption is higher not only in the new steady state but also along the entire transition path. When the capital stock exceeds the Golden Rule level, reducing saving is clearly a good policy, for it increases consumption at every point in time.

Starting with Too Little Capital

When the economy begins with less capital than in the Golden Rule steady state, the policymaker must raise the saving rate to reach the Golden Rule. Figure 8-9 shows what happens. The increase in the saving rate at time t_0 causes an immediate fall in consumption and a rise in investment. Over time, higher investment causes the capital stock to rise. As capital accumulates, output, consumption, and investment gradually increase, approaching the new steady-state levels. Because the initial steady state was below the Golden Rule, the increase in saving eventually leads to a higher level of consumption than that which prevailed initially.

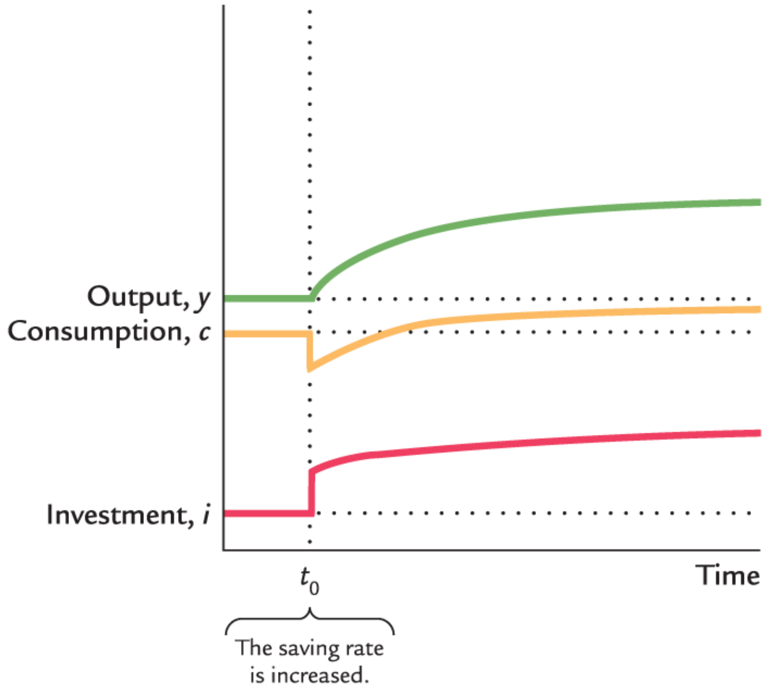


FIGURE 8-9 Increasing Saving When Starting with Less Capital Than in the Golden Rule Steady State This figure shows what happens over time to output, consumption, and investment when the economy begins with less capital than the Golden Rule level and the saving rate is increased. The increase in the saving rate (at time t0t_0) causes an immediate drop in consumption and an equal jump in investment. Over time, as the capital stock grows, output, consumption, and investment increase together. Because the economy began with less capital than the Golden Rule level, the new steady state has a higher level of consumption than the initial steady state.

Does the increase in saving that leads to the Golden Rule steady state raise economic welfare? Eventually it does because the new steady-state level of consumption is higher than the initial level. But achieving that new steady state requires an initial period of reduced consumption. Note the contrast to the case in which the economy begins above the Golden Rule. *When the economy begins above the Golden Rule, reaching the*

Golden Rule produces higher consumption at all points in time. When the economy begins below the Golden Rule, reaching the Golden Rule requires initially reducing consumption to increase consumption in the future.

When deciding whether to try to reach the Golden Rule steady state, policymakers have to take into account that current consumers and future consumers are not always the same people. Reaching the Golden Rule achieves the highest steady-state level of consumption and thus benefits future generations. But when the economy is initially below the Golden Rule, reaching the Golden Rule requires raising investment and thus lowering the consumption of current generations. Thus, when choosing whether to increase capital accumulation, the policymaker faces a trade-off in the welfare of different generations. A policymaker who cares more about current generations than about future ones may decide not to pursue policies to reach the Golden Rule steady state. By contrast, a policymaker who cares about all generations equally will choose to reach the Golden Rule. Even though current generations will consume less, an infinite number of future generations will benefit by moving to the Golden Rule.

Thus, optimal capital accumulation depends crucially on how we weigh the interests of current and future generations. The biblical Golden Rule tells us, “Do unto others as you would have them do unto you.” If we heed this advice, we give all generations equal weight. In this case, it is optimal to reach the Golden Rule level of capital—which is why it is called the “Golden Rule.”

8-3 Population Growth

The Solow model shows that capital accumulation, by itself, cannot explain sustained economic growth: high rates of saving lead to high growth temporarily, but the economy eventually approaches a steady state in which capital and output are constant. To explain the sustained economic growth that we observe in most parts of the world, we must expand the Solow model to incorporate the other two sources of growth—population growth and technological progress. In this section we add population growth to the model.

Instead of assuming that the population is fixed, as we did in Sections 8-1 and 8-2, we now suppose that the population and the labor force grow at a constant rate *n*. For example, the U.S. population grows about 1 percent per year, so $n = 0.01$. This means that if 150 million people are working one year, then 151.5 million (1.01×150) are working the next year, 153.015 million (1.01×151.5) the year after that, and so on.

The Steady State with Population Growth

How does population growth affect the steady state? To answer this question, we must discuss how population growth, along with investment and depreciation, influences the accumulation of capital per worker. As before, investment raises the capital stock, and depreciation reduces it. But now there is a third force acting to change the amount of capital per worker: the growth in the number of workers causes capital per worker to fall.

We continue to let lowercase letters stand for quantities per worker. Thus, $k = K/L$ is capital per worker, and $y = Y/L$ is output per worker. Remember, however, that the number of workers is growing over time.

The change in the capital stock per worker is

$$\Delta k = i - (\delta + n)\,k.$$

This equation shows how investment, depreciation, and population growth influence the per-worker capital stock. Investment increases *k*, whereas depreciation and population growth decrease *k*. We saw this equation earlier in this chapter for the special case of a constant population $(n = 0)$.

We can think of the term $(\delta + n)\ k$ as defining *break-even investment*—the amount of investment necessary to keep the capital stock per worker constant. Break-even investment includes the depreciation of existing capital, which equals δk. It also includes the amount of investment necessary to provide new workers with capital. The amount of investment necessary for this purpose is *nk* because there are *n* new workers for each existing worker and *k* is the amount of capital per worker. The equation shows that population growth reduces the accumulation of capital per worker much the way depreciation does. Depreciation reduces *k* by wearing out the capital stock, whereas population growth reduces *k* by spreading the capital stock more thinly among a larger population of workers.[5]

Our analysis with population growth now proceeds as it did previously. First, we substitute *sf*(*k*) for *i*. The equation can then be written as

$$\Delta k = sf(k) - (\delta + n)\ k.$$

To see what determines the steady-state level of capital per worker, we use Figure 8-10, which extends the analysis of Figure 8-4 to include the effects of population growth. An economy is in a steady state if capital per worker *k* is unchanging. As before, we designate the steady-state value of *k* as k^*. If *k* is less than k^*, investment is greater than break-even investment, so *k* rises. If *k* is greater than k^*, investment is less than break-even investment, so *k* falls.

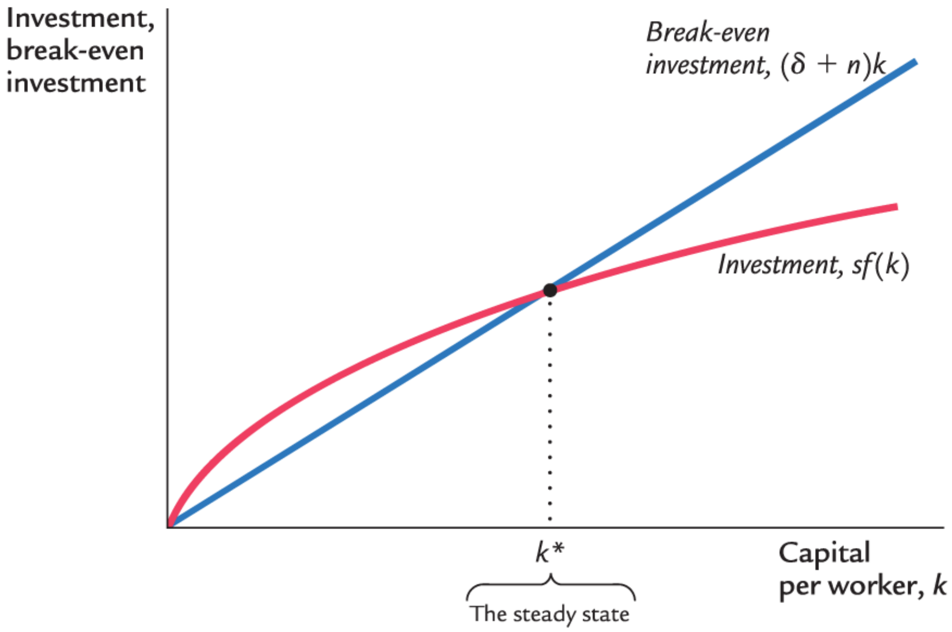


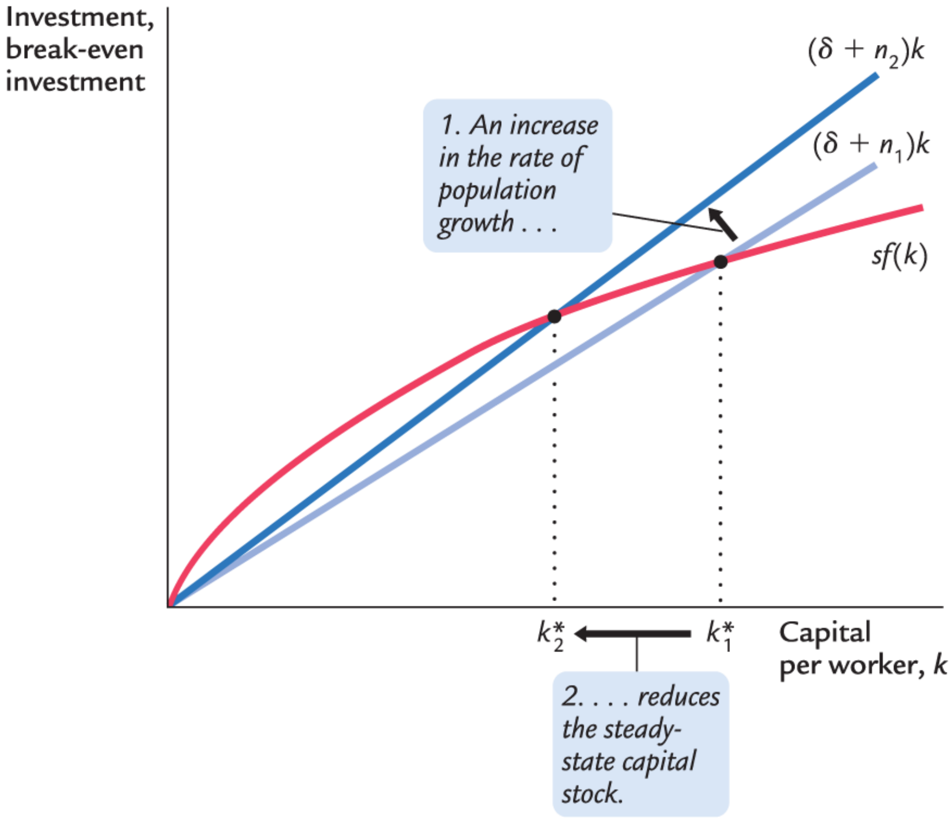
FIGURE 8-10 Population Growth in the Solow Model Depreciation and population growth are two reasons the capital stock per worker shrinks. If *n* is the rate of population growth and δ is the rate of depreciation, then $(\delta + n)\ k$ is *break-even investment*—the amount of investment necessary to keep the capital stock per worker *k* constant. For the economy to be in a steady state, investment *sf*(*k*) must offset the effects of depreciation and population growth $(\delta + n)\ k$. This is represented by the crossing of the two curves.

In the steady state, the positive effect of investment on the capital stock per worker exactly balances the negative effects of depreciation and population growth. That is, at k*, Δk=0k^*, $\Delta k = 0$ and i*=δk*+nk*. $i^* = \delta k^* + nk^*$. Once the economy is in the steady state, investment has two purposes. Some of it (δk*) (δk^*) replaces the depreciated capital, and the rest (nk*)(nk^*) provides the new workers with the steady-state amount of capital.

The Effects of Population Growth

Population growth alters the basic Solow model in three ways. First, it brings us closer to explaining sustained economic growth. In the steady state with population growth, capital per worker and output per worker are constant. Because the number of workers is growing at rate *n*, however, *total* capital and *total* output must also be growing at rate *n*. Hence, although population growth cannot explain sustained growth in the standard of living (because output per worker is constant in the steady state), it can help explain sustained growth in total output.

Second, population growth gives us another reason some countries are rich and others are poor. Consider the effects of an increase in population growth. Figure 8-11 shows that an increase in the rate of population growth from n1n_1 to n2n_2 reduces the steady-state level of capital per worker from k1*k_1^* to k2*.k_2^*. Because k*k^* is lower and because y*=f(k*),$y^* = f(k^*)$, the level of output per worker y*y^* is also lower. Thus, the Solow model predicts that countries with higher population growth will have lower levels of GDP per person. Notice that a change in the population growth rate, like a change in the saving rate, has a level effect on income per person but does not affect the steady-state growth rate of income per person.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 8-11 The Impact of Population Growth An increase in the rate of population growth from n_1 to n_2 shifts the line representing population growth and depreciation upward. The new steady state k_2^* has a lower level of capital per worker than the initial steady state k_1^*. Thus, the Solow model predicts that economies with higher rates of population growth will have lower levels of capital per worker and therefore lower incomes.

Finally, population growth affects our criterion for the Golden Rule (consumption-maximizing) level of capital. To see how this criterion changes, note that consumption per worker is

$$c = y - i.$$

Because steady-state output is $f(k^*)$ and steady-state investment is $(\delta + n)\ k^*$, we can express steady-state consumption as

$$c^* = f(k^*) - (\delta + n)\ k^*.$$

Using an argument largely the same as before, we conclude that the level of k^* that maximizes consumption is the one at which

$$MPK = \delta + n,$$

or, equivalently,

$$MPK - \delta = n.$$

In the Golden Rule steady state, the marginal product of capital net of depreciation equals the rate of population growth.

CASE STUDY

Investment and Population Growth Around the World

We started this chapter with an important question: why are some countries so rich while others are mired in poverty? Our analysis has suggested some answers. According to the Solow model, if a nation devotes a large fraction of its income to saving and investment, it will have a high steady-state capital stock and a high level of income. If a nation saves and invests a small fraction of its income, its steady-state capital and income will be low. In addition, a nation with a high rate of population growth will have a low steady-state capital stock per worker and thus also a low level of income per worker. In other words, high population growth tends to impoverish a country because it is hard to maintain a high level of capital per worker when the number of workers is growing quickly.

To make these points more precisely, recall that in the steady state $\Delta k = 0$, and therefore the steady state is described by the condition

$$sf(k) = (\delta + n)\ k.$$

Now suppose the production function is Cobb–Douglas:

$$y = f(k) = k^{\alpha}.$$

Inverting the production function yields

$$k = y^{1/\alpha}.$$

After substitution for *f*(*k*) and *k*, the steady-state condition can be written

$$sy = (\delta + n)\ y^{1/\alpha}.$$

Solving for y, we obtain

$$y = \left(\frac{s}{\delta+n}\right)^{\alpha/(1-\alpha)}.$$

This equation shows that steady-state income y is positively related to rate of saving and investment s and negatively related to the rate of population growth n. We can think of the variable $s/(\delta + n)$ as measuring the *effective investment rate*. It takes into account not only the percent of income that is saved and invested but also the fact that more investment is needed when capital depreciates and the population grows.

Let's now look at some data to see if this theoretical result helps explain the large international variation in standards of living. Figure 8-12 is a scatterplot using data from about 160 countries. (The figure includes most of the world's economies. It excludes countries whose major source of income is oil, such as Kuwait and Saudi Arabia, because their growth experiences are explained by their unusual circumstances.) On the vertical axis is income per person in 2014. On the horizontal axis is the effective investment rate $s/(\delta + n)$, where s is the average share of investment in GDP and n is the rate of population growth over the preceding 20 years. The depreciation rate δ is assumed to be the same in all countries and is set at 5 percent. The figure shows a strong positive relationship between the effective investment rate $s/(\delta + n)$ and the level of income per person. Thus, the data are consistent with the Solow model's prediction that investment and population growth are key determinants of whether a country is rich or poor.

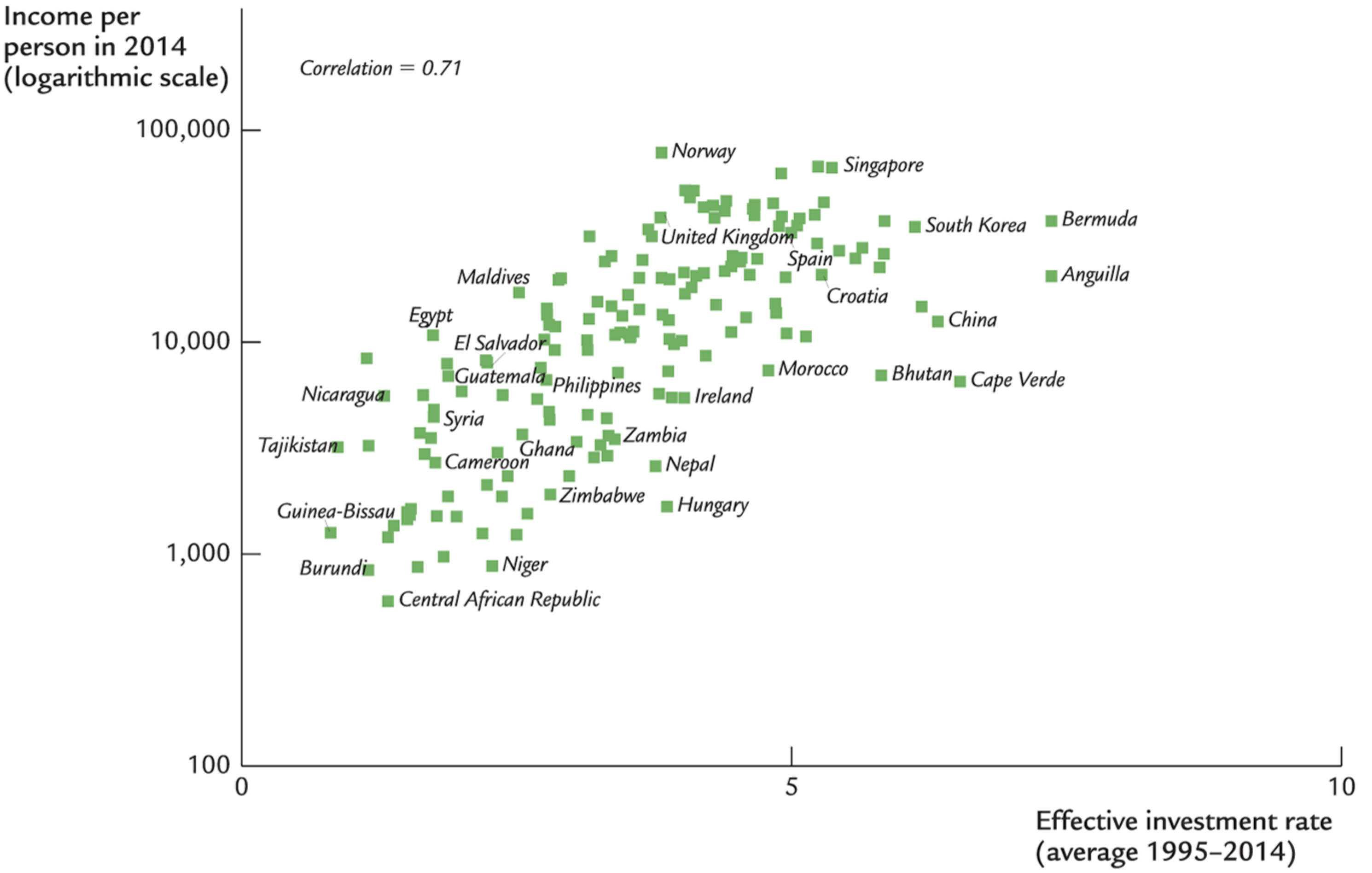


FIGURE 8-12 International Evidence on the Solow Model This scatterplot shows the experience of about 160 countries, each represented by a single point. The vertical axis shows a country's income per person, and the horizontal axis shows its effective investment rate $s/(\delta+n)$.

$s/(\delta+n)$. These two variables are positively associated, as the Solow model predicts.

Data from: Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer, Penn World Table Version 9.0, The Center for International Data at the University of California, Davis and Groningen Growth and Development Centre at the University of Groningen, October 2015.

The positive correlation shown in this figure is an important fact, but it raises as many questions as it resolves. One might naturally ask, for instance, why rates of saving and investment vary from country to country. There are many possible answers, such as tax policy, retirement patterns, the development of financial markets, and cultural differences. In addition, political stability may play a role: not surprisingly, rates of saving and investment tend to be low in countries with frequent wars, revolutions, and coups. Saving and investment also tend to be low in countries with poorly functioning political institutions, as measured by estimates of official corruption.

A final interpretation is reverse causation. Perhaps high levels of income somehow foster high rates of saving and investment. Similarly, high income may reduce population growth, perhaps because birth-control techniques are more readily available in richer countries. The international data can help us evaluate a theory of growth, such as the Solow model, because they show us whether the theory's predictions are borne out in the world. But often more than one theory can explain the same facts. ■

Alternative Perspectives on Population Growth

The Solow model highlights the interaction between population growth and capital accumulation. In this model, high population growth reduces output per worker because rapid growth in the number of workers forces the capital stock to be spread more thinly, so in the steady state, each worker is equipped with less capital. The model omits some other potential effects of population growth. Here we consider two—one emphasizing the interaction of population with natural resources, the other emphasizing the interaction of population with technology.

The Malthusian Model

In his book *An Essay on the Principle of Population as It Affects the Future Improvement of Society*, the early economist Thomas Robert Malthus (1766–1834) offered what may be history's most chilling forecast. Malthus argued that an ever-increasing population would continually strain society's ability to provide for itself. Mankind, he predicted, would forever live in poverty.

Malthus began by noting that "food is necessary to the existence of man" and that "the passion between the sexes is necessary and will remain nearly in its present state." He concluded that "the power of population is infinitely greater than the power in the earth to produce subsistence for man." According to Malthus, the only check on population growth was "misery and vice." Attempts by charities or governments to alleviate poverty were counterproductive, he argued, because they merely allowed the poor to have more children, placing even greater strains on society's productive capabilities.

The Malthusian model may have described the world when Malthus lived, but its prediction that mankind would remain in poverty forever has proven very wrong. The world population has increased about sevenfold over the past two centuries, but average living standards are much higher. Because of economic growth, chronic hunger and malnutrition are less common now than they were in Malthus's day. Famines occur from time to time, but they are more often the result of unequal income distribution or political instability than the inadequate production of food.

Malthus failed to foresee that growth in mankind's ingenuity would more than offset the effects of a larger population. Pesticides, fertilizers, mechanized farm equipment, new crop varieties, and other technological advances that Malthus never imagined have allowed each farmer to feed ever-greater numbers of people. Even with more mouths to feed, fewer farmers are necessary because each farmer is so productive. Today, only about 1 percent of Americans work on farms, producing enough food to feed the nation and some excess to export as well.

In addition, although the "passion between the sexes" is as strong now as it was in Malthus's day, modern birth control has broken the link between passion and population growth. Many advanced nations, such as those in western Europe, are now experiencing fertility below replacement rates. Over the next century, shrinking populations may be more likely than rapidly expanding ones. There is now little reason to think that an ever-expanding population will overwhelm food production and doom mankind to poverty.[6]

The Kremerian Model

While Malthus saw population growth as a threat to rising living standards, economist Michael Kremer has suggested that world population growth is a key driver of advancing prosperity. If there are more people, Kremer argues, there are more scientists, inventors, and engineers to contribute to innovation and technological progress.

As evidence for this hypothesis, Kremer begins by noting that over the broad span of human history, world growth rates have increased together with world population. For example, world growth was more rapid when the world population was 1 billion (which occurred around the year 1800) than it was when the population was only 100 million (around 500 B.C.). This fact is consistent with the hypothesis that having more people induces more technological progress.

Kremer's second, more compelling piece of evidence comes from comparing regions of the world. The melting of the polar ice caps at the end of the ice age around 10,000 B.C. flooded the land bridges and separated the world into several distinct regions that could not communicate with one another for thousands of years. If technological progress is more rapid when there are more people to discover things, then the more populous regions should have experienced more rapid growth.

And, indeed, they did. The most successful region of the world in 1500 (when Columbus reestablished technological contact) included the “Old World” civilizations of the large Eurasia–Africa region. Next in technological development were the Aztec and Mayan civilizations in the Americas, followed by the hunter-gatherers of Australia, and then the primitive people of Tasmania, who lacked even fire-making and most stone and bone tools. The least populous isolated region was Flinders Island, a tiny island between Tasmania and Australia. With few people to contribute new innovations, Flinders Island had the least technological advance and, in fact, seemed to regress. Around 3000 B.C., human society on Flinders Island died out completely.

Kremer concludes from this evidence that a large population is a prerequisite for technological advance.[7]

8-4 Conclusion

This chapter has started building the Solow growth model. The model as developed so far shows how saving and population growth determine the economy's steady-state capital stock and its steady-state level of income per person. It sheds light on many features of actual growth experiences—why Germany and Japan grew so rapidly after being devastated by World War II, why countries that save and invest a high fraction of their output are richer than countries that save and invest a smaller fraction, and why countries with high rates of population growth are poorer than countries with low rates of population growth.

The model cannot, however, explain the persistent growth in living standards we observe in most countries. In the model as it is now, output per worker stops growing when the economy reaches its steady state. To explain persistent growth, we need to introduce technological progress into the model. That is our first job in the next chapter.

CHAPTER 9

Economic Growth II: Technology, Empirics, and Policy

> Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, what, exactly? If not, what is it about the "nature of India" that makes it so? The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else.
>
> —Robert E. Lucas, Jr.

The quotation that opens this chapter was written in 1988. Since then, India has grown rapidly, a phenomenon that has pulled millions of people out of extreme poverty. At the same time, some other poor nations, including many in sub-Saharan Africa, have experienced little growth, and their citizens continue to live meager existences. It is the job of growth theory to explain such disparate outcomes. The reasons that some nations succeed while others fail at promoting long-run economic growth are not easily apparent, but as Robert Lucas suggests, the consequences for human welfare are indeed staggering.

This chapter continues our analysis of the forces governing long-run growth. With the basic version of the Solow model as our starting point, we take on four new tasks.

Our first task is to make the Solow model more general and realistic. In Chapter 3 we saw that capital, labor, and technology are the key determinants of a nation's production of goods and services. In Chapter 8 we developed the Solow model to show how changes in capital (through saving and investment) and changes in the labor force (through population growth) affect the economy's output. We are now ready to add the third source of growth—changes in technology—to the mix. The Solow model does not explain technological progress but, instead, takes it as exogenously given and shows how it interacts with other variables in the process of economic growth.

Our second task is to move from theory to empirics. That is, we consider how well the Solow model fits the facts. Over the past few decades, a large literature has examined the predictions of the Solow model and other models of economic growth. It turns out that the glass is both half full and half empty. The Solow model can shed much light on international growth experiences, but it is far from the last word on the subject.

Our third task is to examine how a nation's public policies can influence the level and growth of its citizens' standard of living. In particular, we address six questions: Should our society save more or less? How can policy influence the rate of saving? Are there some types of investment that policy should especially encourage? What institutions ensure that the economy's resources are put to their best use? Can cultural change spur growth? How can policy increase the rate of technological progress? The Solow growth model provides the theoretical framework within which we consider these policy issues.

Our fourth and final task is to consider what the Solow model leaves out. As we have discussed previously, models help us understand the world by simplifying it. After completing an analysis of a model, therefore, it is important to consider whether we have oversimplified matters. In the last section, we examine a new set of theories, called *endogenous growth theories*, which help to explain the technological progress that the Solow model takes as exogenous.

9-1 Technological Progress in the Solow Model

So far, our presentation of the Solow model has assumed an unchanging relationship between the inputs of capital and labor and the output of goods and services. Yet the model can be modified to include exogenous technological progress, which over time expands society's production capabilities.

The Efficiency of Labor

To incorporate technological progress, we must return to the production function that relates total capital *K* and total labor *L* to total output *Y*. Thus far, the production function has been

$$Y = F\left(K,\ L\right).$$

We now write the production function as

$$Y = F\left(K,\ L \times E\right),$$

where *E* is a new (and somewhat abstract) variable called the **efficiency of labor**. The efficiency of labor is meant to reflect society's knowledge about production methods: as the available technology improves, the efficiency of labor rises, and each hour of work contributes more to the production of goods and services. For instance, the efficiency of labor rose when assembly-line production transformed manufacturing in the early twentieth century, and it rose again when computerization was introduced in the late twentieth century. The efficiency of labor also rises when there are improvements in the health, education, or skills of the labor force.

The term $L \times E$ can be interpreted as measuring the *effective number of workers*. It takes into account the number of actual workers *L* and the efficiency of each worker *E*. In other words, *L* measures the number of workers in the labor force, whereas $L \times E$ measures both the workers and the technology with which the typical worker comes equipped. This new production function states that total output *Y* depends on the inputs of capital *K* and effective workers $L \times E$.

The essence of this approach to modeling technological progress is that increases in the efficiency of labor

E are analogous to increases in the labor force *L*. Suppose, for example, that an advance in production methods makes the efficiency of labor *E* double between 1980 and 2015. This means that a single worker in 2015 is, *in effect*, as productive as two workers were in 1980. That is, even if the actual number of workers (*L*) stays the same from 1980 to 2015, the effective number of workers $(L \times E)$ doubles, and the economy benefits from the increased production of goods and services.

The simplest assumption about technological progress is that it causes the efficiency of labor *E* to grow at some constant rate *g*. For example, if $g = 0.02,$ then each unit of labor becomes 2 percent more efficient each year: output increases as if the labor force had increased by 2 percent more than it really did. This form of technological progress is called *labor augmenting*, and *g* is called the rate of **labor-augmenting technological progress**. Because the labor force *L* is growing at rate *n*, and the efficiency of each unit of labor *E* is growing at rate *g*, the effective number of workers $L \times E$ is growing at rate $n + g.$

The Steady State with Technological Progress

Because technological progress is modeled here as labor augmenting, it fits into the model in much the same way as population growth. Technological progress does not cause the actual number of workers to increase, but because each worker in effect comes with more units of labor over time, technological progress causes the effective number of workers to increase. Thus, the analytic tools we used in Chapter 8 to study the Solow model with population growth are easily adapted to studying the Solow model with labor-augmenting technological progress.

We begin by reconsidering our notation. Previously, before we added technological progress, we analyzed the economy in terms of quantities per worker; now we can generalize that approach by analyzing the economy in terms of quantities per effective worker. We now let $k = K/(L \times E)$ stand for capital per effective worker and $y = Y/(L \times E)$ stand for output per effective worker. With these definitions, we can again write $y = f(k).$

Our analysis proceeds just as it did when we examined population growth. The equation showing the evolution of *k* over time becomes

$$\Delta k = sf(k) - (\delta + n + g)k.$$

As before, the change in the capital stock Δk equals investment *sf*(*k*) minus break-even investment $(\delta + n + g)k.$ Now, however, because $k = K/(L \times E),$ break-even investment includes three terms: to keep *k* constant, δk is needed to replace depreciating capital, *nk* is needed to provide capital for new workers, and *gk* is needed to provide capital for the new "effective workers" created by technological progress.[1]

As shown in Figure 9-1, the inclusion of technological progress does not substantially alter our analysis of the steady state. There is one level of *k*, denoted k^*, at which capital per effective worker and output per effective worker are constant. As before, this steady state represents the long-run equilibrium.

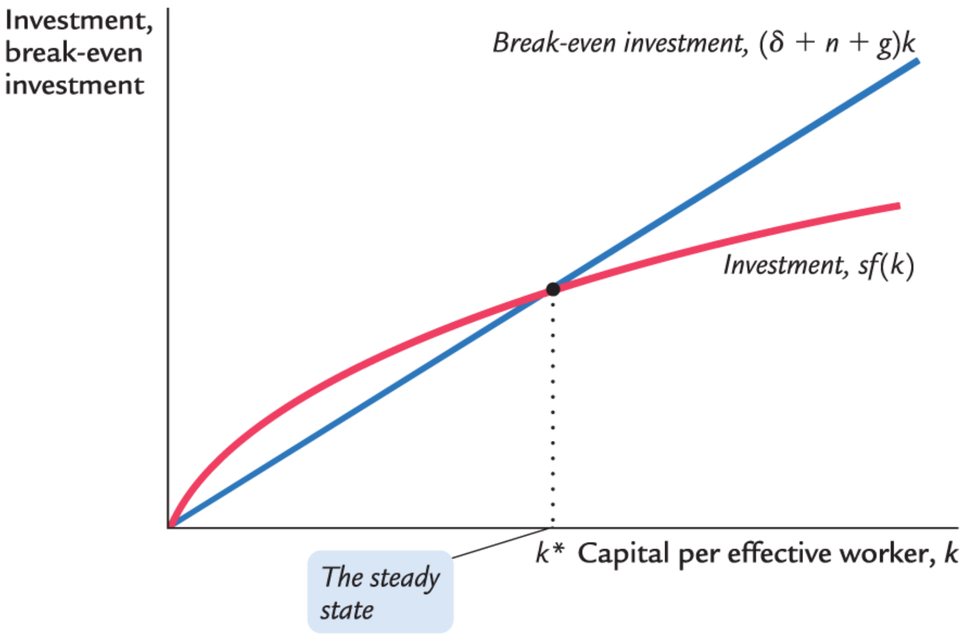


FIGURE 9-1 Technological Progress and the Solow Growth Model Labor-augmenting technological progress at rate *g* enters our analysis of the Solow growth model in much the same way as did population growth at rate *n*. Now that *k* is defined as the amount of capital per effective worker, increases in the effective number of workers because of technological progress tend to decrease *k*. In the steady state, investment *sf* (*k*) exactly offsets the reductions in *k* attributable to depreciation, population growth, and technological progress.

The Effects of Technological Progress

Table 9-1 shows how four key variables behave in the steady state with technological progress. As we have just seen, capital per effective worker *k* is constant in the steady state. Because $y = f(k)$, output per effective worker is also constant. It is these quantities per effective worker that are steady in the steady state.

TABLE 9-1 Steady-State Growth Rates in the Solow Model with Technological Progress

Variable	Symbol	Steady-State Growth Rate
Capital per effective worker	$k=K/(E\times L)$	0
Output per effective worker	$y=Y/(E\times L)=f(k)$	0

Output per worker	Y/L=y×E	*g*
Total output	Y=y×(E×L)	n+g

From this information, we can also infer what is happening to variables that are not expressed in units per effective worker. For instance, consider output per actual worker Y/L=y×E.$Y/L = y \times E$. Because *y* is constant in the steady state and *E* is growing at rate *g*, output per worker must also be growing at rate *g* in the steady state. Similarly, the economy's total output is Y=y×(E×L).$Y = y \times (E \times L)$. Because *y* is constant in the steady state, *E* is growing at rate *g*, and *L* is growing at rate *n*, total output grows at rate n+g$n + g$ in the steady state.

With the addition of technological progress, our model can finally explain the sustained increases in standards of living that we observe. That is, we have shown that technological progress can lead to sustained growth in output per worker. By contrast, a high rate of saving leads to a high rate of growth only until the steady state is reached. Once the economy is in steady state, the rate of growth of output per worker depends only on the rate of technological progress. *According to the Solow model, only technological progress can explain sustained growth and persistently rising living standards*.

The introduction of technological progress also modifies the criterion for the Golden Rule. The Golden Rule level of capital is now defined as the steady state that maximizes consumption per effective worker. Following the same arguments that we have used before, we can show that steady-state consumption per effective worker is

c*=f (k*)−(δ+n+g)k*.$$c^* = f\ (k^*) - (\delta + n + g)\, k^*.$$

Steady-state consumption is maximized if

MPK=δ+n+g,$$MPK = \delta + n + g,$$

or

MPK−δ=n+g.$$MPK - \delta = n + g.$$

That is, at the Golden Rule level of capital, the net marginal product of capital MPK−δ$MPK - \delta$ equals the rate of growth of total output n+g.$n + g$. Because actual economies experience both population growth and technological progress, we must use this criterion to evaluate whether they have more or less capital than they would at the Golden Rule steady state.

9-2 From Growth Theory to Growth Empirics

So far in this chapter we have introduced exogenous technological progress into the Solow model to explain sustained growth in standards of living. Let's now discuss what happens when this theory is forced to confront the facts.

Balanced Growth

According to the Solow model, technological progress causes the values of many variables to rise together in the steady state. This property, called *balanced growth*, does a good job of describing the long-run data for the U.S. economy.

Consider first output per worker Y/L and the capital stock per worker K/L. According to the Solow model, in the steady state both variables grow at g, the rate of technological progress. U.S. data for the past half-century show that output per worker and the capital stock per worker have in fact grown at approximately the same rate—about 2 percent per year. In other words, the capital–output ratio has remained approximately constant over time.

Technological progress also affects factor prices. Problem 4(d) at the end of this chapter asks you to show that, in the steady state, the real wage grows at the rate of technological progress. The real rental price of capital, however, is constant over time. Again, these predictions hold true for the United States. Over the past 50 years, the real wage has increased about 2 percent per year; it has increased at about the same rate as real GDP per worker. Yet the real rental price of capital (measured as real capital income divided by the capital stock) has remained about the same.

The Solow model's prediction about factor prices—and the success of this prediction—is especially noteworthy when contrasted with Karl Marx's theory of the development of capitalist economies. Marx predicted that the return to capital would decline over time and that this would lead to economic and political crisis. Economic history has not supported Marx's prediction, which partly explains why we now study Solow's theory of growth rather than Marx's.

Convergence

If you travel around the world, you will see vast differences in living standards. Income per person in the United States is about eleven times that in Pakistan. Income in Germany is about eight times that in Nigeria. These income disparities are reflected in most measures of the quality of life—from the prevalence of TVs, cell phones, and Internet access to clean water availability, infant mortality, and life expectancy.

Much research has been devoted to the question of whether economies move toward one another over time. That is, do economies that start off poor subsequently grow faster than economies that start off rich? If they do, then the world's poor economies will tend to catch up with the world's rich economies. This process of catch-up is called *convergence*. If convergence does not occur, then countries that start off behind are likely to remain poor.

The Solow model predicts when convergence should occur. According to the model, whether two economies will converge depends on why they differ in the first place. On the one hand, suppose two economies happen by historical accident to start off with different capital stocks, but they have the same steady state, as determined by their saving rates, population growth rates, and efficiency of labor. In this case, we should expect the two economies to converge; the poorer economy with the smaller capital stock will naturally grow more quickly to reach the steady state. (In Chapter 8, we applied this logic to explain rapid growth in Germany and Japan after World War II.) On the other hand, if two economies have different steady states, perhaps because the economies have different rates of saving, then we should not expect convergence. Instead, each economy will approach its own steady state.

Experience is consistent with this analysis. In samples of economies with similar cultures and policies, studies find that economies converge to one another at a rate of about 2 percent per year. That is, the gap between rich and poor economies closes by about 2 percent each year. An example is the economies of individual American states. For historical reasons, such as the Civil War of the 1860s, income levels varied greatly among states at the end of the nineteenth century. Yet these differences have slowly disappeared over time. This convergence can be explained with the Solow model under the assumption that those state economies had different starting points but are approaching a common steady state.

In international data, a more complex picture emerges. When researchers examine only data on income per person, they find little evidence of convergence: countries that start off poor do not grow faster on average than countries that start off rich. This finding suggests that different countries have different steady states. If statistical techniques are used to control for some of the determinants of the steady state, such as saving rates, population growth rates, and accumulation of human capital (education), then once again the data show convergence at a rate of about 2 percent per year. In other words, the economies of the world exhibit *conditional convergence*: they appear to be converging to their own steady states, which in turn are determined by such variables as saving, population growth, and human capital.[2]

Factor Accumulation Versus Production Efficiency

As a matter of accounting, international differences in income per person can be attributed to either differences in the factors of production, such as the quantities of physical and human capital, or differences in the efficiency with which economies use their factors of production. That is, a worker in a poor country may be poor because she lacks tools and skills or because the tools and skills she has are not being put to their best use. To describe this issue in terms of the Solow model, the question is whether the large gap between rich and poor is explained by (1) differences in capital accumulation (including human capital) or (2) differences in the production function.

Much research has attempted to estimate the relative importance of these two sources of income disparities. The exact answer varies from study to study, but both factor accumulation and production efficiency appear to be important. Moreover, a common finding is that they are positively correlated: nations with high levels of physical and human capital also tend to use those factors efficiently.[3]

There are several ways to interpret this positive correlation. One hypothesis is that an efficient economy may encourage capital accumulation. For example, a person in a well-functioning economy may have greater resources and incentive to stay in school and accumulate human capital. Another hypothesis is that capital accumulation may induce greater efficiency. If there are positive externalities to physical and human capital, then countries that save and invest more will appear to have better production functions (unless the research study accounts for these externalities, which is hard to do). Thus, greater production efficiency may cause greater factor accumulation—or the other way around.

A final hypothesis is that both factor accumulation and production efficiency are driven by a common third variable. Perhaps the common third variable is the quality of the nation's institutions, including the government's policymaking process. As one economist put it, when governments screw up, they screw up big time. Bad policies, such as high inflation, excessive budget deficits, widespread market interference, and rampant corruption, often go hand in hand. We should not be surprised that economies exhibiting these maladies both accumulate less capital and fail to use the capital they have as efficiently as they might.

CASE STUDY

Good Management as a Source of Productivity

Incomes vary around the world in part because some nations have higher production efficiency than others. A similar phenomenon is observed within nations: some firms exhibit greater production efficiency than others. Why might that be?

One possible answer is management practices. Some firms are well run; others less so. A well-run firm uses state-of-the-art operations, monitors the performance of its workers, sets challenging but reasonable targets for performance, and provides incentives for workers to put forth their best efforts. Good management means that a firm is getting the most it can from the factors of production it uses.

An influential study by Nicholas Bloom and John Van Reenen documents the importance of good management, as well as some of the reasons that not all firms have it. Bloom and Van Reenen began by surveying 732 medium-sized manufacturing firms in four nations: France, Germany, the United Kingdom, and the United States. They asked various questions about how firms were managed and then graded each firm on how well it conformed to best practices. For example, a firm that promoted employees based on performance was graded higher than one that promoted employees based on how long they had been at the firm.

Perhaps not surprisingly, Bloom and Van Reenen found substantial heterogeneity in the quality of management. In each country, some firms were well run and some were badly run. More noteworthy is that the distribution of management quality differed substantially across the four nations. Firms in the United States had the highest average grade, followed by Germany, then France, and finally the United Kingdom. Much of the cross-country variation came from the prevalence of especially badly run firms: firms with the lowest management grades were much more common in the United Kingdom and France than in the United States and Germany.

The study's next finding is that these management grades were correlated with measures of firm performance. Holding other things equal (such as the size of the firm's capital stock and workforce), well-managed firms had more sales, greater profits, higher stock market values, and lower bankruptcy rates.

If good management leads to all these desirable outcomes, why don't all firms adopt the best practices? Bloom and Van Reenen offer two explanations for the persistence of bad management.

The first is the absence of competition. When a firm with poor management practices is shielded from vigorous competition, its managers can take the easy path and muddle through. By contrast, when a firm operates in a highly competitive market, bad management tends to lead to losses, which eventually induce the firm to either change its practices or close its doors. As a result, in competitive markets, only firms with good management survive. One determinant of competition is openness to trade: when firms have to compete with similar firms around the world, it is hard to maintain bad management practices.

A second explanation for the persistence of bad management is primogeniture—the tradition of some family-owned firms to appoint as chief executive officer (CEO) the family's eldest son. This practice means that the CEO position may not be going to the person who is most qualified for it. Moreover, if the eldest son knows he will get the job by virtue of birth order, rather than having to compete for it with professional managers or at least other family members, he may have less incentive to put in the effort necessary to become a good manager. Indeed, Bloom and Van Reenen report that firms with eldest sons as CEOs are more likely to obtain poor management grades. They also find that primogeniture is far more common in the United Kingdom and France than it is in the United States and Germany, perhaps because of the long-lasting influence of the Norman tradition.

The bottom line from this study is that differences in management practices can help explain why some nations have higher productivity and thus higher incomes than others. These differences in management, in turn, may be traced to differences in degrees of competition and historical traditions.[4] ■

9-3 Policies to Promote Growth

So far we have used the Solow model to uncover the theoretical relationships among the different sources of economic growth, and we have discussed some of the empirical work that describes actual growth experiences. We can now use the theory and evidence to help guide our thinking about economic policy.

Evaluating the Rate of Saving

According to the Solow growth model, how much a nation saves and invests is a key determinant of its citizens' standard of living. So let's begin our policy discussion with a natural question: Is the rate of saving in the U.S. economy too low, too high, or just right?

As we have seen, the saving rate determines the steady-state levels of capital and output. One particular saving rate produces the Golden Rule steady state, which maximizes consumption per worker and thus economic well-being. The Golden Rule provides the benchmark against which we can compare the U.S. economy.

To decide whether the U.S. economy is at, above, or below the Golden Rule steady state, we need to compare the marginal product of capital net of depreciation (MPK−δ)$(MPK - \delta)$ with the growth rate of total output (n+g).$(n + g)$. As we established in Section 9-1, at the Golden Rule steady state, MPK−δ=n+g.$MPK - \delta = n + g$. If the economy is operating with less capital than in the Golden Rule steady state, then diminishing marginal product tells us that MPK−δ>n+g.$MPK - \delta > n + g$. In this case, increasing the rate of saving will increase capital accumulation and economic growth and, eventually, lead to a steady state with higher consumption (although consumption will be lower for part of the transition to the new steady state). On the other hand, if the economy has more capital than in the Golden Rule steady state, then MPK−δ<n+g.$MPK - \delta < n + g$. In this case, capital accumulation is excessive: reducing the rate of saving will lead to higher consumption both immediately and in the long run.

To make this comparison for a real economy, such as the U.S. economy, we need an estimate of the growth rate of output (n+g)$(n + g)$ and an estimate of the net marginal product of capital (MPK−δ).$(MPK - \delta)$. Real GDP in the United States grows an average of 3 percent per year, so n+g=0.03.$n + g = 0.03$. We can estimate the net marginal product of capital from the following three facts:

1. The capital stock is about 2.5 times one year's GDP.
2. Depreciation of capital is about 10 percent of GDP.

3. Capital income is about 30 percent of GDP.

Using the notation of our model (and the result from Chapter 3 that capital owners earn income of *MPK* for each unit of capital), we can write these facts as

1. $k = 2.5y.$
2. $\delta k = 0.1y.$
3. $MPK \times k = 0.3y.$

We solve for the rate of depreciation δ by dividing equation 2 by equation 1:

$$\begin{aligned} \delta k/k &= (0.1y)\,/\,(2.5y) \\ \delta &= 0.04. \end{aligned}$$

And we solve for the marginal product of capital *MPK* by dividing equation 3 by equation 1:

$$\begin{aligned} (MPK \times k)/k &= (0.3y)/(2.5y) \\ MPK &= 0.12. \end{aligned}$$

Thus, about 4 percent of the capital stock depreciates each year, and the marginal product of capital is about 12 percent per year. The net marginal product of capital $MPK - \delta$ is about 8 percent per year.

We can now see that the return to capital ($MPK - \delta = 8$ percent per year) is well above the economy's average growth rate ($n + g = 3$ percent per year). This fact, together with our previous analysis, indicates that the capital stock in the U.S. economy is well below the Golden Rule level. In other words, if the United States saved and invested a higher fraction of its income, it would grow more rapidly and eventually reach a steady state with higher consumption.

This conclusion is not unique to the U.S. economy. When similar calculations are done for other economies, the results are much the same. The possibility of excessive saving and capital accumulation beyond the Golden Rule level is intriguing as a matter of theory, but it appears not to be a problem that actual economies face. In practice, economists are more often concerned with insufficient saving. It is this kind of calculation that provides the intellectual foundation for this concern.[5]

Changing the Rate of Saving

The preceding calculations show that to move the U.S. economy toward the Golden Rule steady state, policymakers should enact policies to encourage national saving. But how can they do that? We saw in Chapter 3 that, as a matter of simple accounting, higher national saving means higher public saving, higher private saving, or some combination of the two. Much of the debate over policies to increase growth centers on which of these options is likely to be most effective.

The most direct way in which the government affects national saving is through public saving—the difference between what the government receives in tax revenue and what it spends. When its spending exceeds its revenue, the government runs a *budget deficit*, which represents negative public saving. As we saw in Chapter 3, a budget deficit raises interest rates and crowds out investment; the resulting reduction in the capital stock is part of the burden of the national debt on future generations. Conversely, if it spends less than it raises in revenue, the government runs a *budget surplus*, which it can use to retire some of the national debt and stimulate investment.

The government also affects national saving by influencing private saving—the saving done by households and firms. How much people decide to save depends on the incentives they face, and these incentives are altered by various public policies. Many economists argue that high tax rates on capital—including the corporate income tax, the federal income tax, the estate tax, and many state income and estate taxes—discourage private saving by reducing the rate of return that savers earn. On the other hand, tax-exempt retirement accounts, such as IRAs, are designed to encourage private saving by giving preferential treatment to income saved in these accounts. Some economists have proposed increasing the incentive to save by replacing the current system of income taxation with a system of consumption taxation.

Many disagreements over public policy are rooted in different views about how much private saving responds to incentives. For example, suppose the government increased the amount that people can put into tax-exempt retirement accounts. Would people respond to this incentive by saving more? Or, instead, would people merely transfer saving already done in taxable savings accounts into these tax-advantaged accounts, reducing tax revenue and thus public saving without any stimulus to private saving? The desirability of the policy depends on the answers to these questions. Unfortunately, despite much research on this issue, no consensus has emerged.

Allocating the Economy’s Investment

The Solow model makes the simplifying assumption that there is only one type of capital. In the world, of course, there are many types. Private businesses invest in traditional types of capital, such as bulldozers and steel plants, and newer types of capital, such as computers and robots. The government invests in various forms of public capital, called *infrastructure*, such as roads, bridges, and sewer systems.

In addition, there is *human capital*—the knowledge and skills that workers acquire through education, from early childhood programs such as Head Start to on-the-job training for adults in the labor force. Although the capital variable in the Solow model is usually interpreted as including only physical capital, in many ways human capital is analogous to physical capital. Like physical capital, human capital increases our ability to produce goods and services. Raising the level of human capital requires investment in the form of teachers, libraries, and student time. Research on economic growth has emphasized that human capital is at least as important as physical capital in explaining international differences in standards of living. One way of modeling this fact is to give the variable we call "capital" a broader definition that includes both human and physical capital.[6]

Policymakers trying to promote economic growth must confront the issue of what kinds of capital the economy needs most. In other words, what kinds of capital yield the highest marginal products? To a large extent, policymakers can rely on the marketplace to allocate the pool of saving to alternative types of investment. Those industries with the highest marginal products of capital will naturally be most willing to borrow at market interest rates to finance new investment. Many economists advocate that the government should merely create a "level playing field" for different types of capital—for example, by ensuring that the tax system treats all forms of capital equally. The government can then rely on the market to allocate capital efficiently.

Other economists have suggested that the government should promote specific forms of capital. Suppose, for instance, that technological advance occurs as a byproduct of certain activities. This would happen if new and improved production processes are devised during the process of building capital (a phenomenon called *learning by doing*) and if these ideas become part of society's pool of knowledge. Such a byproduct is called a *technological externality* (or a *knowledge spillover*). In the presence of such externalities, the social returns to capital exceed the private returns, and the benefits of capital accumulation to society are greater than the Solow model suggests.[7] Moreover, some types of capital accumulation may yield greater externalities than others. If, for example, installing robots yields greater technological externalities than building a new steel mill, then perhaps the government should use the tax laws to encourage investment in robots. The success of such an *industrial policy*, as it is sometimes called, depends on the government's ability to accurately measure the externalities of different economic activities so that it can give the correct incentives.

Most economists are skeptical about industrial policies for two reasons. First, measuring the externalities from different sectors is hard. If policy is based on poor measurements, its effects might be close to random and, thus, worse than no policy at all. Second, the political process is far from perfect. Once the government gets into the business of rewarding specific industries with subsidies and tax breaks, the rewards are as likely to be based on political clout as on the magnitude of externalities.

One type of capital that necessarily involves the government is public capital. Local, state, and federal governments are always deciding if and when they should borrow to finance new roads, bridges, and transit systems. In 2016, for example, President Donald Trump was elected promising a $1 trillion increase in infrastructure spending. Among economists, this proposal had both defenders and critics. Yet all of them agree that measuring the marginal product of public capital is difficult. Private capital generates an easily measured rate of profit for the firm owning the capital, whereas the benefits of public capital are more diffuse. Furthermore, while private capital investment is made by investors spending their own money, the allocation of resources for public capital involves the political process and taxpayer funding. It is all too common to see “bridges to nowhere” being built simply because the local senator or member of Congress has the political muscle to get funds approved.

CASE STUDY

Industrial Policy in Practice

Policymakers and economists have long debated whether the government should promote certain industries and firms because they are strategically important for the economy. In the United States, the debate goes back over two centuries. Alexander Hamilton, the first U.S. Secretary of the Treasury, favored tariffs on certain imports to encourage the development of domestic manufacturing. The Tariff of 1789 was the second act passed by the new federal government. The tariff helped manufacturers, but it hurt farmers, who had to pay more for foreign-made products. Because the North was home to most of the manufacturers, while the South had more farmers, the tariff was one source of the regional tensions that eventually led to the Civil War.

Advocates of a significant government role in promoting technology can point to some successes. For example, the precursor of the modern Internet is a system called ARPANET, which was established by the U.S. Department of Defense as a way for information to flow among military installations. There is little doubt that the Internet has been associated with large advances in productivity and that the government had a hand in its creation. According to proponents of industrial policy, this example illustrates how the government can help jump-start an emerging technology.

Yet governments can also make mistakes when they try to supplant private business decisions. Japan’s Ministry of International Trade and Industry (MITI) is sometimes viewed as a successful practitioner of industrial policy, but it once tried to stop Honda from expanding its business from motorcycles to automobiles. MITI thought that the nation already had enough car manufacturers. Fortunately, the government lost this battle, and Honda turned into one of the world’s largest and most profitable car companies. Soichiro Honda, the company’s founder, once said, “Probably I would have been even more successful had we not had MITI.”

More recently, government policy has aimed to promote “green technologies.” In particular, the U.S. federal government has subsidized the production of energy in ways that yield lower carbon emissions, which contribute to global climate change. It is too early to judge the long-run success of this policy, but there have been some short-run embarrassments. In 2011, a manufacturer of solar panels called Solyndra declared bankruptcy just two years after the federal government granted it a $535 million loan guarantee.

The debate over industrial policy will surely continue in the years to come. The final judgment about this kind of government intervention in the market requires evaluating both the efficiency of unfettered markets and the ability of governmental institutions to identify technologies worthy of support. ■

Establishing the Right Institutions

As we discussed earlier, economists who study international differences in the standard of living attribute some of these differences to the inputs of physical and human capital and some to the productivity with which these inputs are used. One reason nations may have different levels of production efficiency is that they have different institutions guiding the allocation of scarce resources. Creating the right institutions is important for ensuring that resources are allocated to their best use.

Jason Reed/Reuters/Newscom

North and South Korea from space.

Source: Reuters.

Perhaps the clearest current example of the importance of institutions is the comparison between North and South Korea. For many centuries, these two nations were combined with a common government, heritage, culture, and economy. Yet in the aftermath of World War II, an agreement between the United States and the Soviet Union split Korea in two. Above the thirty-eighth parallel, North Korea established institutions based on the Soviet model of authoritarian communism. Below the thirty-eighth parallel, South Korea established institutions based on the American model of democratic capitalism. Today, the difference in economic development could not be more stark. GDP per person in North Korea is less than one-tenth of what it is in South Korea. This difference is visible in satellite photos taken at night. South Korea is well lit—its widespread use of electricity a sign of advanced economic development. North Korea, in contrast, is shrouded in darkness.

Among democratic capitalist nations, there are important but more subtle institutional differences. One example is a nation's legal tradition. Some countries, such as the United States, Australia, India, and Singapore, are former colonies of the United Kingdom and, therefore, have English-style common-law systems. Other nations, such as Italy, Spain, and most of those in Latin America, have legal traditions that evolved from the French Napoleonic Code. Studies have found that legal protections for shareholders and creditors are stronger in English-style than French-style legal systems. As a result, the English-style countries have better-developed capital markets. Nations with better-developed capital markets, in turn, experience more rapid growth because it is easier for small and start-up companies to finance investment projects, leading to a more efficient allocation of the nation's capital.[8]

Another important institutional difference across countries is the quality of government and honesty of government officials. Ideally, governments should provide a "helping hand" to the market system by protecting property rights, enforcing contracts, promoting competition, prosecuting fraud, and so on. Yet governments can diverge from this ideal and act more like a "grabbing hand" by using the authority of the state to enrich the powerful at the expense of the broader community. Empirical studies have shown that the extent of corruption in a nation is indeed a significant determinant of economic growth.[9]

Adam Smith, the great eighteenth-century economist, was well aware of the role of institutions in economic growth. He once wrote, "Little else is requisite to carry a state to the highest degree of opulence from the lowest barbarism but peace, easy taxes, and a tolerable administration of justice: all the rest being brought about by the natural course of things." Sadly, many nations do not enjoy these three simple advantages.

CASE STUDY

The Colonial Origins of Modern Institutions

International data show a remarkable correlation between latitude and economic prosperity: nations closer to the equator typically have lower levels of income per person than nations farther from the equator. This fact is true in both the Northern and Southern Hemispheres.

What explains the correlation? Some economists have suggested that the tropical climates near the equator have a direct negative impact on productivity. In the heat of the tropics, agriculture is more difficult, and disease is more prevalent. This makes the production of goods and services more difficult.

Although the direct impact of geography is one reason tropical nations tend to be poor, it is not the whole story. Research by Daron Acemoglu, Simon Johnson, and James Robinson has suggested an indirect mechanism—the impact of geography on institutions. Here is their explanation, presented in several steps:

1. In the seventeenth, eighteenth, and nineteenth centuries, tropical climates presented European settlers with an increased risk of disease, especially malaria and yellow fever. As a result, when Europeans were colonizing much of the rest of the world, they avoided settling in tropical areas, such as most of Africa and Central America. The European settlers preferred areas with more moderate climates and better health conditions, such as the regions that are now the United States, Canada, and New Zealand.
2. In those areas where Europeans settled in large numbers, the settlers established European-style institutions that protected property rights and limited the power of government. By contrast, in tropical climates, the colonial powers often set up "extractive" institutions, including authoritarian governments, so they could take advantage of the area's natural resources. These institutions enriched the colonizers, but

they did little to foster economic growth.

3. Although the era of colonial rule is now long over, the early institutions that the European colonizers established are strongly correlated with the modern institutions in the former colonies. In tropical nations, where the colonial powers set up extractive institutions, there is typically less protection of property rights even today. When the colonizers left, the extractive institutions remained and were simply taken over by new ruling elites.
4. The quality of institutions is a key determinant of economic performance. Where property rights are well protected, people have more incentive to make the investments that lead to economic growth. Where property rights are less respected, as is typically the case in tropical nations, investment and growth tend to lag behind.

This research suggests that much of the international variation in living standards that we observe today is a result of the long reach of history.[10] ■

Supporting a Pro-growth Culture

A nation's *culture* refers to the values, attitudes, and beliefs of its people. Many social scientists have suggested that culture can have an important influence on economic growth. For example, in his classic 1905 book *The Protestant Ethic and the Spirit of Capitalism*, sociologist Max Weber argued that the acceleration of economic growth in northern Europe beginning in the sixteenth century can be attributed to the rise of Calvinism, a branch of Protestantism that emphasizes hard work and frugality.

Culture has many facets and is hard to quantify. Yet there are some clear ways in which cultural differences can help explain why some nations are rich and others are poor. Here are four examples:

- Societies differ in their treatment of women. In some nations, prevailing cultural norms keep women poorly educated and out of the labor force, depressing the standard of living.
- Societies differ in their attitudes toward children—both how many to have and how much to educate them. Higher population growth can depress incomes, and greater human capital can increase it.
- Societies differ in how open they are to new ideas, especially ideas from abroad. More open nations can quickly adopt technological advances wherever they occur, while less open ones find themselves further from the world's technological frontier.
- Societies differ in how much people trust one another. Because the legal system is a costly and imperfect mechanism for enforcing agreements, it is easier to coordinate economic activities when trust is high. Indeed, there is a positive correlation between the level of trust as reported in surveys and a nation's income per person. Trust is related to what some economists call *social capital*, the network of cooperative relationships among people, including such diverse groups as churches and bowling leagues.

A nation's culture arises from various historical, anthropological, and sociological forces and is not easily controlled by policymakers. But culture evolves over time, and policy can play a supporting role. The changing attitude toward women in the United States over the past century is a case in point. Women today get more education and are more likely to be in the labor force than they were in the past, and these changes have led to a higher standard of living for American families. Public policy was not the main cause of these developments, but laws expanding educational opportunities for women and protecting women's rights in the workplace were complementary with the evolution of culture.

Encouraging Technological Progress

The Solow model shows that sustained growth in income per worker must come from technological progress. The Solow model, however, takes technological progress as exogenous; it does not explain it. Unfortunately, the determinants of technological progress are not well understood.

Despite this limited understanding, many public policies are designed to stimulate technological progress. Most of these policies encourage the private sector to devote resources to technological innovation. For example, the patent system gives a temporary monopoly to inventors of new products; the tax code offers tax breaks for firms engaging in research and development; and government agencies, such as the National Science Foundation, subsidize basic research in universities. In addition, as discussed previously, proponents of industrial policy argue that the government should take a more active role in promoting specific industries that are key to rapid technological advance.

In recent years, the encouragement of technological progress has taken on an international dimension. Many of the companies that engage in research to advance technology are located in the United States and other developed nations. Developing nations such as China have an incentive to "free ride" on this research by not strictly enforcing intellectual property rights. That is, Chinese companies often use ideas developed abroad without compensating the patent holders. The United States has objected to this practice, and China has promised to step up enforcement. If intellectual property rights were better enforced around the world, firms would have more incentive to engage in research, and this would promote worldwide technological progress.

CASE STUDY

Is Free Trade Good for Economic Growth?

At least since Adam Smith, economists have advocated free trade as a policy that promotes national prosperity. Here is how Smith put the argument in his 1776 classic, *The Wealth of Nations*:

> It is a maxim of every prudent master of a family, never to attempt to make at home what it will cost him more to make than to buy. The tailor does not attempt to make his own shoes, but buys them of the shoemaker. The shoemaker does not attempt to make his own clothes but employs a tailor. . . .
>
> What is prudence in the conduct of every private family can scarce be folly in that of a great kingdom. 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.

Today, economists make the case with greater rigor, relying on David Ricardo's theory of comparative advantage as well as more modern theories of international trade. According to these theories, a nation open to trade can achieve greater production efficiency and a higher standard of living by specializing in those goods for which it has a comparative advantage.

A skeptic might point out that this is just a theory. What about the evidence? Do nations that permit free trade in fact enjoy greater prosperity? A large body of literature addresses precisely this question.

One approach is to look at international data to see if countries that are open to trade typically enjoy greater prosperity. The evidence shows that they do. Economists Andrew Warner and Jeffrey Sachs studied this question for the period from 1970 to 1989. They report that among developed nations, the open economies grew at 2.3 percent per year, while the closed economies grew at 0.7 percent per year. Among developing nations, the open economies grew at 4.5 percent per year, while the closed economies again grew at 0.7 percent per year. These findings are consistent with Smith's view that trade enhances prosperity, but they are not conclusive. Correlation does not prove causation. Perhaps being closed to trade is correlated with various other restrictive government policies, and it is those other policies that retard growth.

A second approach is to look at what happens when closed economies remove their trade restrictions. Once again, Smith's hypothesis fares well. Throughout history, when nations open themselves up to the world economy, the typical result is a subsequent increase in economic growth. This occurred in Japan in the 1850s, South Korea in the 1960s, and Vietnam in the 1990s. But once again, correlation does not prove causation. Trade liberalization is often accompanied by other reforms aimed to promote growth, and it is hard to disentangle the effects of trade from the effects of the other reforms.

A third approach to measuring the impact of trade on growth, proposed by economists Jeffrey Frankel and David Romer, is to look at the impact of geography. Some countries trade less simply because they are geographically disadvantaged. For example, New Zealand is disadvantaged compared to Belgium because it is farther from other populous countries. Similarly, landlocked countries are disadvantaged compared to countries with their own seaports. Because these geographic characteristics are correlated with trade, but arguably uncorrelated with other determinants of prosperity, they can be used to identify the causal impact of trade on income. (The statistical technique, which you may have studied in an econometrics course, is called *instrumental variables*.) After analyzing the data, Frankel and Romer conclude that "a rise of one percentage point in the ratio of trade to GDP increases income per person by at least one-half percent. Trade appears to raise income by spurring the accumulation of human and physical capital and by increasing output for given levels of capital."

The overwhelming weight of the evidence from this body of research is that Adam Smith was right: openness to international trade is good for economic growth.[11] ■

9-4 Beyond the Solow Model: Endogenous Growth Theory

> A chemist, a physicist, and an economist are trapped on a desert island, trying to figure out how to open a can of food.
>
> “Let’s heat the can over the fire until it explodes,” says the chemist.
>
> “No, no,” says the physicist, “let’s drop the can onto the rocks from the top of a high tree.”
>
> “I have an idea,” says the economist. “First, we assume a can opener . . .”

This old joke takes aim at how economists use assumptions to simplify—and sometimes oversimplify—the problems they face. It is particularly apt when evaluating the theory of economic growth. One goal of growth theory is to explain the persistent rise in living standards that we observe in most parts of the world. The Solow growth model shows that such persistent growth must come from technological progress. But where does technological progress come from? In the Solow model, it is just assumed!

To fully understand the process of economic growth, we need to go beyond the Solow model and develop models that explain technological advance. Models that do this often go by the label **endogenous growth theory** because they reject the Solow model’s assumption of exogenous technological change. Although the field of endogenous growth theory is large and sometimes complex, here we get a quick taste of this modern research.[12]

The Basic Model

To illustrate the idea behind endogenous growth theory, let’s start with a particularly simple production function:

$$Y = AK,$$

where *Y* is output, *K* is the capital stock, and *A* is a constant that measures the amount of output produced for each unit of capital. Notice that this production function does not exhibit the property of diminishing returns to capital. One extra unit of capital produces *A* extra units of output, regardless of how much capital there is. This absence of diminishing returns to capital is the key difference between this endogenous growth model and the Solow model.

Now let's see what this production function says about economic growth. As before, we assume a fraction *s* of income is saved and invested. We therefore describe capital accumulation with an equation similar to those we used previously:

$$\Delta K = sY - \delta K.$$

This equation states that the change in the capital stock ΔK equals investment *sY* minus depreciation δK. Combining this equation with the $Y = AK$ production function, we obtain, after a bit of manipulation,

$$\Delta Y / Y = \Delta K / K = sA - \delta.$$

This equation shows what determines the growth rate of output $\Delta Y / Y$. Notice that if $sA > \delta$, the economy's income grows forever, even without the assumption of exogenous technological progress.

Thus, a simple change in the production function can dramatically alter the predictions about economic growth. In the Solow model, saving temporarily leads to growth, but diminishing returns to capital eventually force the economy to approach a steady state in which growth depends only on exogenous technological progress. By contrast, in this endogenous growth model, saving and investment can lead to persistent growth.

But is it reasonable to abandon the assumption of diminishing returns to capital? The answer depends on how we interpret the variable *K* in the production function $Y = AK$. If we take the traditional view that *K* includes only the economy's stock of plants and equipment, then it is natural to assume diminishing returns. Giving 10 computers to a worker does not make that worker 10 times as productive as she is with one computer.

Advocates of endogenous growth theory, however, argue that the assumption of constant (rather than diminishing) returns to capital is more palatable if *K* is interpreted more broadly. Perhaps the best case can be made for the endogenous growth model by viewing knowledge as a type of capital. Clearly, knowledge is a key input into the economy's production—both its production of goods and services and its production of new knowledge. Compared to other forms of capital, however, it is less natural to assume that knowledge exhibits the property of diminishing returns. (Indeed, the increasing pace of scientific and technological innovation over the past few centuries has led some economists to argue that there are increasing returns to knowledge.) If we accept the view that knowledge is a type of capital, then this endogenous growth model with its assumption of constant returns to capital becomes a more plausible description of long-run economic growth.

A Two-Sector Model

Although the $Y = AK$ model is the simplest example of endogenous growth, the theory has gone well beyond this. One line of research has tried to develop models with more than one sector of production in order to offer a better description of the forces that govern technological progress. To see what we might learn from such models, let's sketch out an example.

The economy has two sectors, which we can call manufacturing firms and research universities. Firms produce goods and services, which are used for consumption and investment in physical capital. Universities produce a factor of production called "knowledge," which is then freely used in both sectors. The economy is described by the production function for firms, the production function for universities, and the capital-accumulation equation:

$$\begin{aligned} Y &= F[K,\ (1-u)LE] && \text{(production function in manufacturing firms)}, \\ \Delta E &= g(u)E && \text{(production function in research universities)}, \\ \Delta K &= sY - \delta K && \text{(capital accumulation)}, \end{aligned}$$

where u is the fraction of the labor force in universities (and $1-u$ is the fraction in manufacturing), E is the stock of knowledge (which in turn determines the efficiency of labor), and g is a function that shows how the growth in knowledge depends on the fraction of the labor force in universities. The rest of the notation is standard. As usual, the production function for the manufacturing firms is assumed to have constant returns to scale: if we double both the amount of physical capital (K) and the effective number of workers in manufacturing $[(1-u)\,LE]$, we double the output of goods and services (Y).

This model is a cousin of the $Y = AK$ model. Most important, this economy exhibits constant (rather than diminishing) returns to capital, as long as capital is broadly defined to include knowledge. In particular, if we double both physical capital K and knowledge E, then we double the output of both sectors in the economy. As a result, like the $Y = AK$ model, this model can generate persistent growth without the assumption of exogenous shifts in the production function. Here persistent growth arises endogenously because the creation of knowledge in universities never slows down.

At the same time, however, this model is also a cousin of the Solow growth model. If u, the fraction of the labor force in universities, is held constant, then the efficiency of labor E grows at the constant rate $g(u)$. This result of constant growth in the efficiency of labor at rate g is precisely the assumption made in the Solow model with technological progress. Moreover, the rest of the model—the manufacturing production function and the capital-accumulation equation—also resembles the rest of the Solow model. As a result, for any given value of u, this endogenous growth model works just like the Solow model.

There are two key decision variables in this model. As in the Solow model, the fraction of output used for saving and investment, s, determines the steady-state stock of physical capital. In addition, the fraction of labor in universities, u, determines the growth in the stock of knowledge. Both s and u affect the level of income, although only u affects the steady-state growth rate of income. Thus, this model of endogenous growth takes a small step in the direction of showing which societal decisions determine the rate of technological change.

The Microeconomics of Research and Development

The two-sector endogenous growth model just presented takes us closer to understanding technological progress, but it still tells only a rudimentary story about the creation of knowledge. If one thinks about the process of research and development for even a moment, three facts become apparent. First, although knowledge is largely a public good (that is, a good freely available to everyone), much research is done in firms that are driven by the profit motive. Second, research is profitable because innovations give firms temporary monopolies, either because of the patent system or because there is an advantage to being the first firm on the market with a new product. Third, when one firm innovates, other firms build on that innovation to produce the next generation of innovations. These (essentially microeconomic) facts are not easily connected with the (essentially macroeconomic) growth models we have discussed so far.

Some endogenous growth models try to incorporate these facts about research and development. Doing this requires modeling both the decisions that firms face as they engage in research and the interactions among firms that have some degree of monopoly power over their innovations. Going into more detail about these models is beyond the scope of this book, but it should be clear already that one virtue of these endogenous growth models is that they offer a more complete description of the process of technological innovation.

One question these models are designed to address is whether, from the standpoint of society as a whole, private profit-maximizing firms tend to engage in too little or too much research. In other words, is the social return to research (which is what society cares about) greater or smaller than the private return (which is what motivates individual firms)? As a theoretical matter, there are effects in both directions. On the one hand, when a firm creates a new technology, it makes other firms better off by giving them a base of knowledge on which to build in future research. As Isaac Newton famously remarked, “If I have seen further, it is by standing on the shoulders of giants." On the other hand, when one firm invests in research, it can also make other firms worse off if it does little more than become the first to discover a technology that another firm would have invented in due course. This duplication of research effort has been called the "stepping on toes" effect. Whether firms left to their own devices do too little or too much research depends on whether the positive "standing on shoulders" externality or the negative "stepping on toes" externality is more prevalent.

Although theory alone is ambiguous about whether research effort is more or less than optimal, the empirical work in this area is usually less so. Many studies have suggested the "standing on shoulders" externality is important and, as a result, the social return to research is large—often more than 40 percent per year. This is an impressive rate of return, especially when compared to the return to physical capital, which we earlier estimated to be about 8 percent per year. In the judgment of some economists, this finding justifies substantial government subsidies to research.[13]

The Process of Creative Destruction

In his 1942 book *Capitalism, Socialism, and Democracy*, economist Joseph Schumpeter suggested that economic progress comes through a process of **creative destruction**. According to Schumpeter, the driving force behind progress is the entrepreneur with an idea for a new product, a new way to produce an old product, or some other innovation. When the entrepreneur's firm enters the market, it has some degree of monopoly power over its innovation; indeed, it is the prospect of monopoly profits that motivates the entrepreneur. The entry of the new firm is good for consumers, who have an expanded range of choices, but it is often bad for incumbent producers, who have to compete with the entrant. If the new product is sufficiently better than old ones, some incumbents may be driven out of business. Over time, the process keeps renewing itself. The entrepreneur's firm becomes an incumbent, enjoying high profitability until its product is displaced by another entrepreneur with the next generation of innovation.

History confirms Schumpeter's thesis that there are winners and losers from technological progress. For example, in England in the early nineteenth century, an important innovation was the invention and spread of weaving machines that could be operated by unskilled workers, allowing manufacturers to produce textiles at low cost. This technological advance was good for consumers, who could clothe themselves more cheaply. Yet skilled textile artisans in England saw their jobs threatened by the new technology, and they responded by organizing violent revolts. The rioting workers, called Luddites, smashed the machines used in the wool and cotton mills and set the homes of the mill owners on fire (a less than creative form of destruction). Today, the term "Luddite" refers to anyone who opposes technological progress.

A recent example of creative destruction involves the retail giant Walmart. Although retailing may seem like a static activity, in fact it is a sector that has seen sizable rates of technological progress over the past several decades. Through better inventory-control, marketing, and personnel-management techniques, Walmart has found ways to bring goods to consumers at lower cost than traditional retailers. These changes benefit consumers, who can buy goods at lower prices, and the stockholders of Walmart, who share in its profitability. But they adversely affect small mom-and-pop stores, which find it hard to compete when a Walmart opens nearby.

Faced with the prospect of being the victims of creative destruction, incumbent producers often look to the political process to stop the entry of new, more efficient competitors. The original Luddites wanted the British government to save their jobs by restricting the spread of the new textile technology; instead, Parliament sent troops to suppress the Luddite riots. Similarly, in recent years, local retailers have sometimes tried to use local land-use regulations to stop Walmart from entering their market. The cost of such entry restrictions, however, is a slower pace of technological progress. In Europe, where entry regulations are stricter than they are in the United States, the economies have not seen the emergence of retailing giants like Walmart; as a result, productivity growth in retailing has been lower.[14]

Schumpeter's vision of how capitalist economies work has merit as a matter of economic history. Moreover, it has inspired some recent work in the theory of economic growth. One line of endogenous growth theory, pioneered by economists Philippe Aghion and Peter Howitt, builds on Schumpeter's insights by modeling technological advance as a process of entrepreneurial innovation and creative destruction.[15]

9-5 Conclusion

Long-run economic growth is the most important determinant of the economic well-being of a nation's citizens. Everything else that macroeconomists study—unemployment, inflation, trade deficits, and so on—pales in comparison.

Fortunately, economists know quite a lot about the forces that govern economic growth. The Solow growth model and the more recent endogenous growth models show how saving, population growth, and technological progress interact in determining the level and growth of a nation's standard of living. These theories offer no magic recipe to ensure that an economy achieves rapid growth, but they give much insight, and they provide the intellectual framework for much of the debate over public policy aimed at promoting long-run economic growth.

APPENDIX

Accounting for the Sources of Economic Growth

Real GDP in the United States has grown an average of about 3 percent per year over the past 50 years. What explains this growth? In Chapter 3 we linked the output of the economy to the factors of production—capital and labor—and to the production technology. Here we develop a technique called *growth accounting* that divides the growth in output into three different sources: increases in capital, increases in labor, and advances in technology. This breakdown provides us with a measure of the rate of technological change.

Increases in the Factors of Production

We first examine how increases in the factors of production contribute to increases in output. To do this, we start by assuming there is no technological change, so the production function relating output *Y* to capital *K* and labor *L* is constant over time:

$$Y = F(K, L).$$

In this case, the amount of output changes only because the amount of capital or labor changes.

Increases in Capital

First, consider changes in capital. If the amount of capital increases by ΔK units, by how much does the amount of output increase? To answer this question, we need to recall the definition of the marginal product of capital *MPK*:

$$MPK = F(K+1, L) - F(K, L).$$

The marginal product of capital tells us how much output increases when capital increases by 1 unit.

Therefore, when capital increases by ΔK units, output increases by approximately $MPK \times \Delta K$.[16]

For example, suppose the marginal product of capital is 1/5; that is, an additional unit of capital increases the amount of output produced by one-fifth of a unit. If we increase the amount of capital by 10 units, we can compute the amount of additional output as follows:

$$\begin{aligned}\Delta Y &= MPK \times \Delta K \\ &= 1/5\ \frac{\text{units of output}}{\text{unit of capital}} \times 10 \text{ units of capital} \\ &= 2 \text{ units of output.}\end{aligned}$$

By increasing capital by 10 units, we obtain 2 more units of output. Thus, we use the marginal product of capital to convert changes in capital into changes in output.

Increases in Labor

Next, consider changes in labor. If the amount of labor increases by ΔL units, by how much does output increase? We answer this question the same way we answered the question about capital. The marginal product of labor *MPL* tells us how much output changes when labor increases by 1 unit—that is,

$$MPL = F(K, L+1) - F(K, L).$$

Therefore, when the amount of labor increases by ΔL units, output increases by approximately $MPL \times \Delta L$.

For example, suppose the marginal product of labor is 2; that is, an additional unit of labor increases the amount of output produced by 2 units. If we increase the amount of labor by 10 units, we can compute the amount of additional output as follows:

$$\begin{aligned}\Delta Y &= MPL \times \Delta K \\ &= 2\ \frac{\text{units of output}}{\text{unit of labor}} \times 10 \text{ units of labor} \\ &= 20 \text{ units of output.}\end{aligned}$$

By increasing labor by 10 units, we obtain 20 more units of output. Thus, we use the marginal product of labor to convert changes in labor into changes in output.

Increases in Capital and Labor

Finally, let's consider the more realistic case in which both factors of production change. Suppose the amount of capital increases by ΔK and the amount of labor increases by ΔL. The increase in output then comes from two sources: more capital and more labor. We can divide this increase into the two sources, using the marginal products of the two inputs:

$$\Delta Y = (MPK \times \Delta K) + (MPL \times \Delta L).$$

The first term in parentheses is the increase in output resulting from the increase in capital; the second term in parentheses is the increase in output resulting from the increase in labor. This equation shows us how to attribute growth to each factor of production.

We now want to convert this last equation into a form that is easier to interpret and apply to the available data. First, with some algebraic rearrangement, the equation becomes[17]

$$\frac{\Delta Y}{Y} = \left(\frac{MPK \times K}{Y}\right) \frac{\Delta K}{K} + \left(\frac{MPL \times L}{Y}\right) \frac{\Delta L}{L}.$$

This form of the equation relates the growth rate of output $\Delta Y/Y$ to the growth rate of capital $\Delta K/K$ and the growth rate of labor $\Delta L/L$.

Next, we need to find some way to measure the terms in parentheses in the last equation. In Chapter 3 we showed that the marginal product of capital equals its real rental price. Therefore, $MPK \times K$ is the total return to capital, and $(MPK \times K)/Y$ is capital's share of output. Similarly, the marginal product of labor equals the real wage. Therefore, $MPL \times L$ is the total compensation that labor receives, and $(MPL \times L)/Y$ is labor's share of output. Under the assumption that the production function has constant returns to scale, Euler's theorem (which we discussed in Chapter 3) tells us that these two shares sum to 1. In this case, we can write

$$\frac{\Delta Y}{Y} = \alpha \frac{\Delta K}{K} + (1 - \alpha) \frac{\Delta L}{L},$$

where α is capital's share and $(1-\alpha)$ is labor's share.

This last equation gives us a simple formula for showing how changes in inputs lead to changes in output. It shows, in particular, that we must weight the growth rates of the inputs by the factor shares. Capital's share in the United States is about 30 percent—that is, $\alpha = 0.30$. Therefore, a 10 percent increase in the amount of capital $(\Delta K/K = 0.10)$ leads to a 3 percent increase in the amount of output $(\Delta Y/Y = 0.03)$. Similarly, a 10 percent increase in the amount of labor $(\Delta L/L = 0.10)$ leads to a 7 percent increase in the amount of output $(\Delta Y/Y = 0.07)$.

Technological Progress

So far in our analysis of the sources of growth, we have been assuming that the production function does not change over time. In practice, of course, technological progress improves the production function. For any given amount of inputs, we can produce more output today than we could in the past. We now extend the analysis to allow for technological progress.

We include the effects of the changing technology by writing the production function as

$$Y = AF\ (K,\ L),$$

where *A* is a measure of the current level of technology called *total factor productivity*. Output now increases not only because of increases in capital and labor but also because of increases in total factor productivity. If total factor productivity increases by 1 percent and if the inputs are unchanged, then output increases by 1 percent.

Allowing for a changing level of technology adds another term to our equation accounting for economic growth:

$$\begin{array}{ccccccc} \frac{\Delta Y}{Y} & = & \alpha\ \frac{\Delta K}{K} & + & (1-\alpha)\ \frac{\Delta L}{L} & + & \frac{\Delta A}{A} \\ \text{Growth in Output} & = & \text{Contribution of Capital} & + & \text{Contribution of Labor} & + & \text{Growth in Total Factor Productivity.} \end{array}$$

This is the key equation of growth accounting. It identifies and allows us to measure the three sources of growth: changes in the amount of capital, changes in the amount of labor, and changes in total factor productivity.

Because total factor productivity is not directly observable, it is measured indirectly. We have data on the growth in output, capital, and labor; we also have data on capital's share of output. From these data and the growth-accounting equation, we can compute the growth in total factor productivity to make sure everything adds up:

$$\frac{\Delta A}{A} = \frac{\Delta Y}{Y} - \alpha \frac{\Delta K}{K} - (1-\alpha) \frac{\Delta L}{L}.$$

$\Delta A/A$ is the change in output that cannot be explained by changes in inputs. Thus, the growth in total factor productivity is computed as a residual—that is, as the amount of output growth that remains after we have accounted for the determinants of growth that we can measure directly. Indeed, $\Delta A/A$ is sometimes called the *Solow residual,* after Robert Solow, who first showed how to compute it.[18]

Total factor productivity can change for many reasons. Changes most often arise because of increased knowledge about production methods, so the Solow residual is frequently used as a measure of technological progress. Yet other factors, such as education and government regulation, can affect total factor productivity as well. For example, if higher public spending raises the quality of education, then workers may become more productive and output may rise, which implies higher total factor productivity. In another example, if government regulations require firms to purchase capital to reduce pollution or increase worker safety, then the capital stock may rise without any increase in measured output, which implies lower total factor productivity. *Total factor productivity captures anything that changes the relation between measured inputs and measured output.*

The Sources of Growth in the United States

Having learned how to measure the sources of economic growth, we now look at the data. Table 9-2 uses U.S. data to measure the contributions of the three sources of growth between 1948 and 2016.

TABLE 9-2 Accounting for Economic Growth in the United States

			SOURCES OF GROWTH				
Years	**Output Growth** $\Delta Y/Y$	**=**	**Capital** $a\Delta K/K$	**+**	**Labor** $(1-\alpha)\Delta L/L$	+	**Total Factor Productivity** $\Delta A/A$
			(average percentage increase per year)				
1948–2016	3.4		1.3		1.0		1.1

1948–1973	4.3	1.3	1.0	1.9
1973–2016	3.0	1.2	1.0	0.7

Data from: U.S. Department of Labor. Data are for the non-farm business sector. Parts may not add to total due to rounding.

This table shows that output in the non-farm business sector grew an average of 3.4 percent per year during this time. Of this 3.4 percent, 1.3 percent was attributable to increases in the capital stock, 1.0 percent to increases in the labor input, and 1.1 percent to increases in total factor productivity. These data show that increases in capital, labor, and productivity have contributed almost equally to economic growth in the United States.

Table 9-2 also shows that the growth in total factor productivity slowed substantially around 1973. Before 1973, total factor productivity grew at 1.9 percent per year; after 1973, it grew at only 0.7 percent per year. Accumulated over many years, even a small change in the rate of growth has a large effect on economic well-being. Real income in the United States in 2016 would have been about 67 percent higher if productivity growth had remained at its previous level.

CASE STUDY

The Slowdown in Productivity Growth

Why did the slowdown in productivity growth around 1973 occur? There are many hypotheses to explain this adverse phenomenon. Here are four of them.

Measurement Problems One possibility is that the productivity slowdown did not really occur and that it shows up in the data because the data are flawed. As you may recall from Chapter 2, one problem in measuring inflation is correcting for changes in the quality of goods and services. The same issue arises when measuring output and productivity. For instance, if technological advance leads to *more* computers being built, then the increase in output and productivity is easy to measure. But if technological advance leads to *faster* computers being built, then output and productivity have increased, but that increase is more subtle and harder to measure. Government statisticians try to correct for changes in quality, but despite their best efforts, the resulting data are far from perfect.

Unmeasured quality improvements mean that our standard of living is rising more rapidly than the official data indicate. This issue should make us suspicious of the data, but by itself it cannot explain the productivity slowdown. To explain a *slowdown* in growth, one must argue that the measurement problems got *worse*. There is some indication that this might be so. As time passes, fewer people work in industries with tangible and easily measured output, such as agriculture, and more work in industries with intangible and less easily measured output, such as medical services. Yet few economists believe that measurement problems are the full story.

Oil Prices When the productivity slowdown began around 1973, the obvious hypothesis to explain it was the large increase in oil prices caused by the actions of the OPEC oil cartel. The primary piece of evidence was the timing: productivity growth slowed at about the same time that oil prices skyrocketed. Over time, however, this explanation has appeared less likely. One reason is that the accumulated shortfall in productivity seems too large to be explained by an increase in oil prices; petroleum-based products are not that large a fraction of a typical firm's costs. In addition, if this explanation were right, productivity should have sped up when political turmoil in OPEC caused oil prices to plummet in 1986. Unfortunately, that did not happen.

Worker Quality Some economists suggest that the productivity slowdown might have been caused by changes in the labor force. In the early 1970s, the large baby-boom generation started leaving school and taking jobs. At the same time, changing social norms encouraged many women to leave full-time housework and enter the labor force. Both developments lowered the average level of experience among workers, which in turn lowered average productivity.

Other economists point to changes in worker quality, as gauged by human capital. Although the educational attainment of the labor force continued to rise throughout this period, it was not increasing as rapidly as it had in the past. Moreover, declining performance on some standardized tests suggests that the quality of education was declining. If so, this could explain slowing productivity growth.

The Depletion of Ideas Still other economists suggest that in the early 1970s the world started running out of new ideas about how to produce, pushing the economy into an age of slower technological progress. These economists often argue that the anomaly is not in the period since 1970 but in the preceding two decades. In the late 1940s, the economy had a large backlog of ideas that had not been fully implemented because of the Great Depression of the 1930s and World War II in the first half of the 1940s. After the economy used up this backlog, the argument goes, a slowdown in productivity growth was likely. Indeed, although the growth rates after 1973 were disappointing compared to those of the 1950s and 1960s, they were not lower than average growth rates from 1870 to 1950.

Unfortunately, the slowdown in productivity growth remains a mystery. In the middle of the 1990s, productivity growth accelerated, a development often attributed to advances in computers and information technology, but the acceleration proved temporary. For the decade ending in 2016, total factor productivity grew at a mere 0.4 percent per year. The mysterious productivity slowdown remains a feature of the contemporary economy.[19] ■

The Solow Residual in the Short Run

When Robert Solow introduced his famous residual, his aim was to shed light on the forces that determine technological progress and economic growth in the long run. But economist Edward Prescott has looked at the Solow residual as a measure of technological change over shorter periods of time. He concludes that fluctuations in technology are a major source of short-run changes in economic activity.

Figure 9-2 shows the Solow residual and the growth in output using annual data for the United States during the period 1960 to 2016. Notice that the Solow residual fluctuates substantially. If Prescott's interpretation is correct, then we can draw conclusions from these short-run fluctuations, such as that technology worsened in 1982 and improved in 1984. Notice also that the Solow residual moves closely with output: in years when output falls, technology tends to worsen. In Prescott's view, this fact implies that recessions are driven by adverse shocks to technology. The hypothesis that technological shocks are the driving force behind short-run economic fluctuations, and the complementary hypothesis that monetary policy has no role in explaining these fluctuations, is the foundation for an approach called *real-business-cycle theory*.

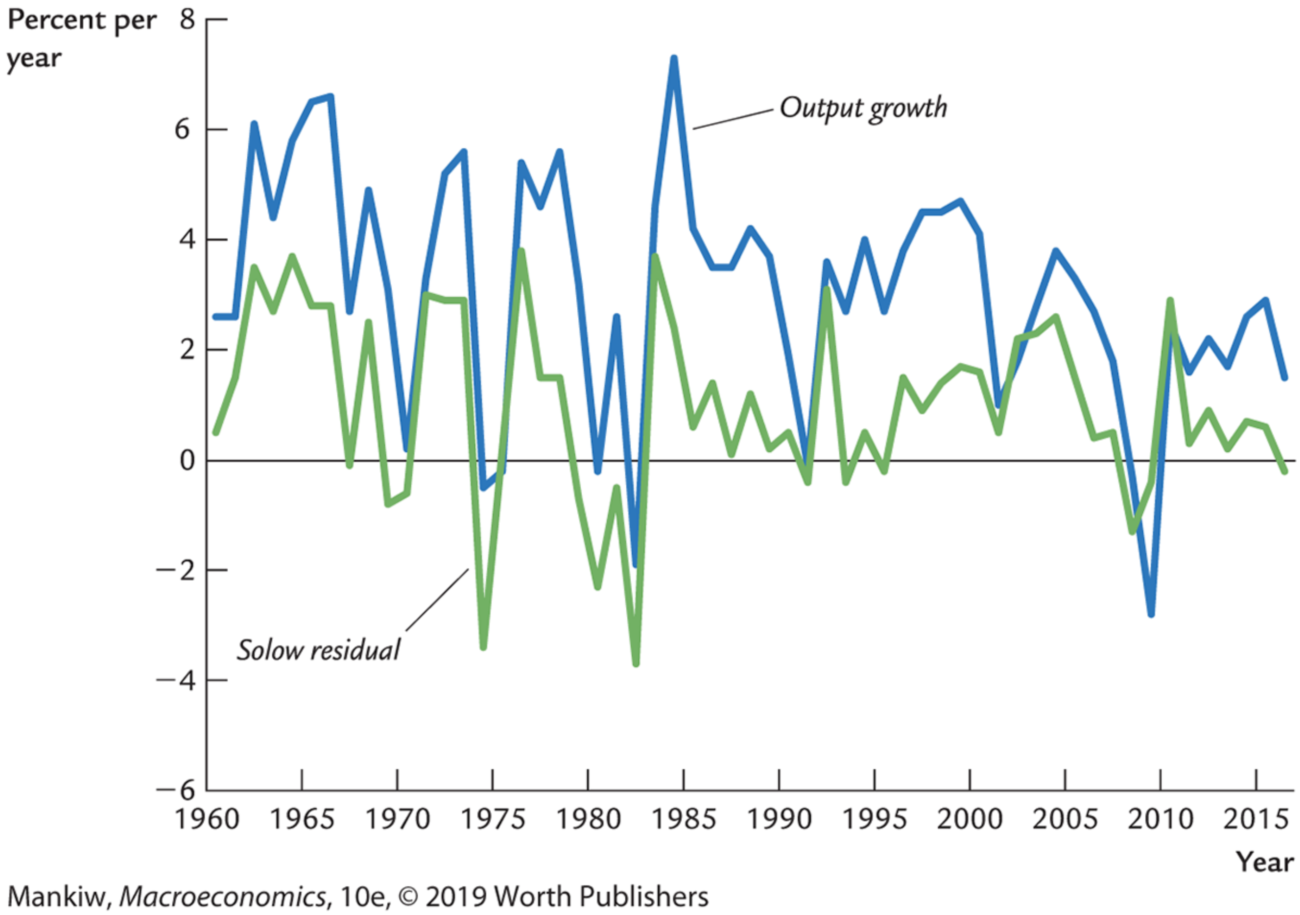


FIGURE 9-2 Growth in Output and the Solow Residual The Solow residual, which some economists interpret as a measure of technology shocks, fluctuates with the economy's output of goods and services.

Data from: U.S. Department of Commerce.

Prescott's interpretation of these data is controversial, however. Many economists believe that the Solow residual does not accurately represent changes in technology over short periods of time. The standard explanation of the cyclical behavior of the Solow residual is that it results from two measurement problems.

First, during recessions, firms may continue to employ workers they do not need so that they will have these workers on hand when the economy recovers. This phenomenon, called *labor hoarding*, means that labor input is overestimated in recessions because the hoarded workers are probably not working as hard as usual. As a result, the Solow residual is more cyclical than the available production technology. In a recession, productivity as measured by the Solow residual falls even if technology has not changed simply because hoarded workers are sitting around waiting for the recession to end.

Second, when demand is low, firms may produce things that are not easily measured. In recessions, workers may clean the factory, organize the inventory, get some training, and do other useful tasks that standard measures of output fail to include. If so, then output is underestimated in recessions, which would also make the measured Solow residual cyclical for reasons other than technology.

Thus, economists can interpret the cyclical behavior of the Solow residual in different ways. Some economists point to the low productivity in recessions as evidence for adverse technology shocks. Others believe that measured productivity is low in recessions because workers are not working as hard as usual and because more of their output is not measured. Unfortunately, there is no clear evidence on the importance of labor hoarding and the cyclical mismeasurement of output. Therefore, different interpretations of Figure 9-2 persist.[20]

MORE PROBLEMS AND APPLICATIONS

1. In the economy of Solovia, the owners of capital get two-thirds of national income, and the workers receive one-third.
   a. The men of Solovia stay at home performing household chores, while the women work in factories. If some of the men started working outside the home so that the labor force increased by 5 percent, what would happen to the measured output of the economy? Does labor productivity—defined as output per worker—increase, decrease, or stay the same? Does total factor productivity increase, decrease, or stay the same?
   b. In year 1, the capital stock was 6, the labor input was 3, and output was 12. In year 2, the capital stock was 7, the labor input was 4, and output was 14. What happened to total factor productivity between the two years?

2. Labor productivity is defined as Y/L, the amount of output divided by the amount of labor input. Start with the growth-accounting equation and show that the growth in labor productivity depends on growth in total factor productivity and growth in the capital–labor ratio. In particular, show that

   $$\frac{\Delta(Y/L)}{Y/L} = \frac{\Delta A}{A} + \alpha \frac{\Delta(K/L)}{K/L}.$$

   (*Hint:* You may find the following mathematical trick helpful.) If $z = wx$, then the growth rate of *z* is approximately the growth rate of *w* plus the growth rate of *x*. That is,

   $$\Delta z/z \approx \Delta w/w + \Delta x/x.$$

3. Suppose an economy described by the Solow model is in a steady state with population growth *n* of 1.8 percent per year and technological progress *g* of 1.8 percent per year. Total output and total capital grow at 3.6 percent per year. Suppose further that the capital share of output is 1/3. If you used the growth-accounting equation to divide output growth into three sources—capital, labor, and total factor productivity—how much would you attribute to each source? Compare your results to the figures for the United States in Table 9-2.

CHAPTER 10

Introduction to Economic Fluctuations

> The modern world regards business cycles much as the ancient Egyptians regarded the overflowing of the Nile. The phenomenon recurs at intervals, it is of great importance to everyone, and natural causes of it are not in sight.
>
> —John Bates Clark, 1898

Economic fluctuations present a recurring problem for economists and policymakers. On average, the real GDP of the United States grows about 3 percent per year. But this long-run average hides the fact that the economy's output of goods and services does not grow smoothly. Growth is higher in some years than in others; sometimes the economy loses ground, and growth turns negative. These fluctuations in output are closely associated with fluctuations in employment. When the economy experiences a period of falling output and rising unemployment, it is said to be in a *recession*.

A recent and severe economic downturn, called the Great Recession, began in late 2007. From the third quarter of 2007 to the third quarter of 2008, the economy's production of goods and services was approximately flat, in contrast to its normal growth. Real GDP then plunged sharply in the fourth quarter of 2008 and first quarter of 2009. The unemployment rate rose from 4.7 percent in November 2007 to 10.0 percent in October 2009. The recession officially ended in June 2009 when growth resumed, but the recovery was weak, and unemployment remained high for several years. The unemployment rate did not fall back below 5 percent until 2016.

Economists call these short-run fluctuations in output and employment the *business cycle*. Although this term suggests that these fluctuations are regular and predictable, they are not. Recessions are as irregular as they are common. Sometimes they occur close together, while at other times, they are farther apart. For example, the United States fell into recession in 1982, only two years after the previous downturn. By the end of that year, the unemployment rate had reached 10.8 percent—the highest level since the Great Depression of the 1930s. But after the 1982 recession, it was eight years before the economy experienced another one.

These historical events raise many questions: What causes short-run fluctuations? What model should we use to explain them? Can policymakers avoid recessions? If so, what policy levers should they use?

In Parts Two and Three of this book, we developed theories to explain how the economy behaves in the long run. Here, in Part Four, we see how economists explain short-run fluctuations. In this chapter, we take on three tasks. First, we examine the data that describe short-run fluctuations. Second, we discuss the key differences between how the economy behaves in the long run and how it behaves in the short run. Third, we introduce the model of aggregate supply and aggregate demand, which most economists use to explain short-run fluctuations. Developing this model in more detail will be our primary job in the chapters that follow.

Just as Egypt now controls the flooding of the Nile Valley with the Aswan Dam, modern society tries to control the business cycle with appropriate economic policies. The model we develop over the next several chapters shows how monetary and fiscal policies influence the business cycle. We will see how these policies can potentially stabilize the economy or, if poorly conducted, make the problem of economic instability even worse.

10-1 The Facts About the Business Cycle

Before discussing the theory of business cycles, let's look at some of the facts that describe short-run fluctuations in economic activity.

GDP and Its Components

The economy's gross domestic product (GDP) measures total income and total expenditure in the economy. Because GDP is the broadest gauge of economic conditions, it is the natural place to start in analyzing the business cycle. Figure 10-1 shows the growth of real GDP from 1970 to 2017. The horizontal line shows the average growth rate of 3 percent per year. You can see that economic growth is not steady and occasionally turns negative.

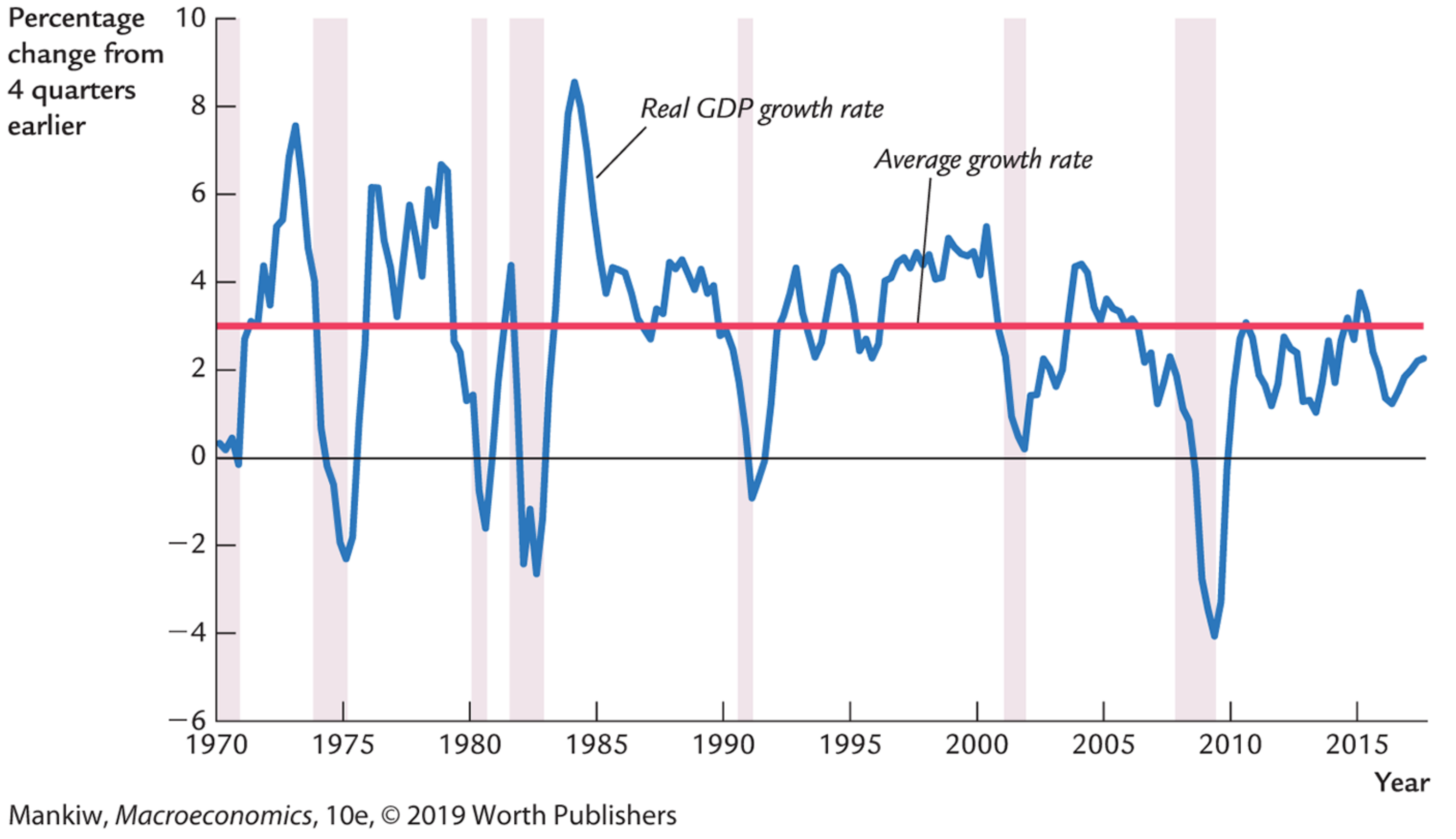


FIGURE 10-1 Real GDP Growth in the United States Growth in real GDP averages about 3 percent per year, but there are substantial fluctuations around this average. The shaded areas represent periods of recession.

Data from: U.S. Department of Commerce, National Bureau of Economic Research.

The shaded areas in the figure indicate periods of recession. The official arbiter of when recessions begin and end is the National Bureau of Economic Research (NBER), a nonprofit economic research group. The

NBER’s Business Cycle Dating Committee (of which the author of this book was once a member) chooses the starting date of each recession, called the business cycle *peak*, and the ending date, called the business cycle *trough*.

“Well, so long Eddie, the recession’s over.”
Mick Stevens/The New Yorker/Conde Nast/The Cartoon Bank

What determines whether a downturn in the economy is sufficiently severe to be deemed a recession? There is no simple answer. According to an old rule of thumb, a recession is a period of at least two consecutive quarters of declining real GDP. This rule, however, does not always hold. For example, the recession of 2001 had two quarters of negative growth, but those quarters were not consecutive. In fact, the NBER’s Business Cycle Dating Committee does not follow any fixed rule but, instead, looks at a variety of data and uses its judgment when picking the starting and ending dates of recessions.[1]

Figure 10-2 shows the growth in two major components of GDP—consumption in panel (a) and investment in panel (b). Growth in both variables declines during recessions. Take note, however, of the scales for the vertical axes. Investment is far more volatile than consumption over the business cycle. When the economy heads into a recession, households respond to the fall in their incomes by consuming less, but the decline in spending on business equipment, structures, new housing, and inventories is even more substantial.

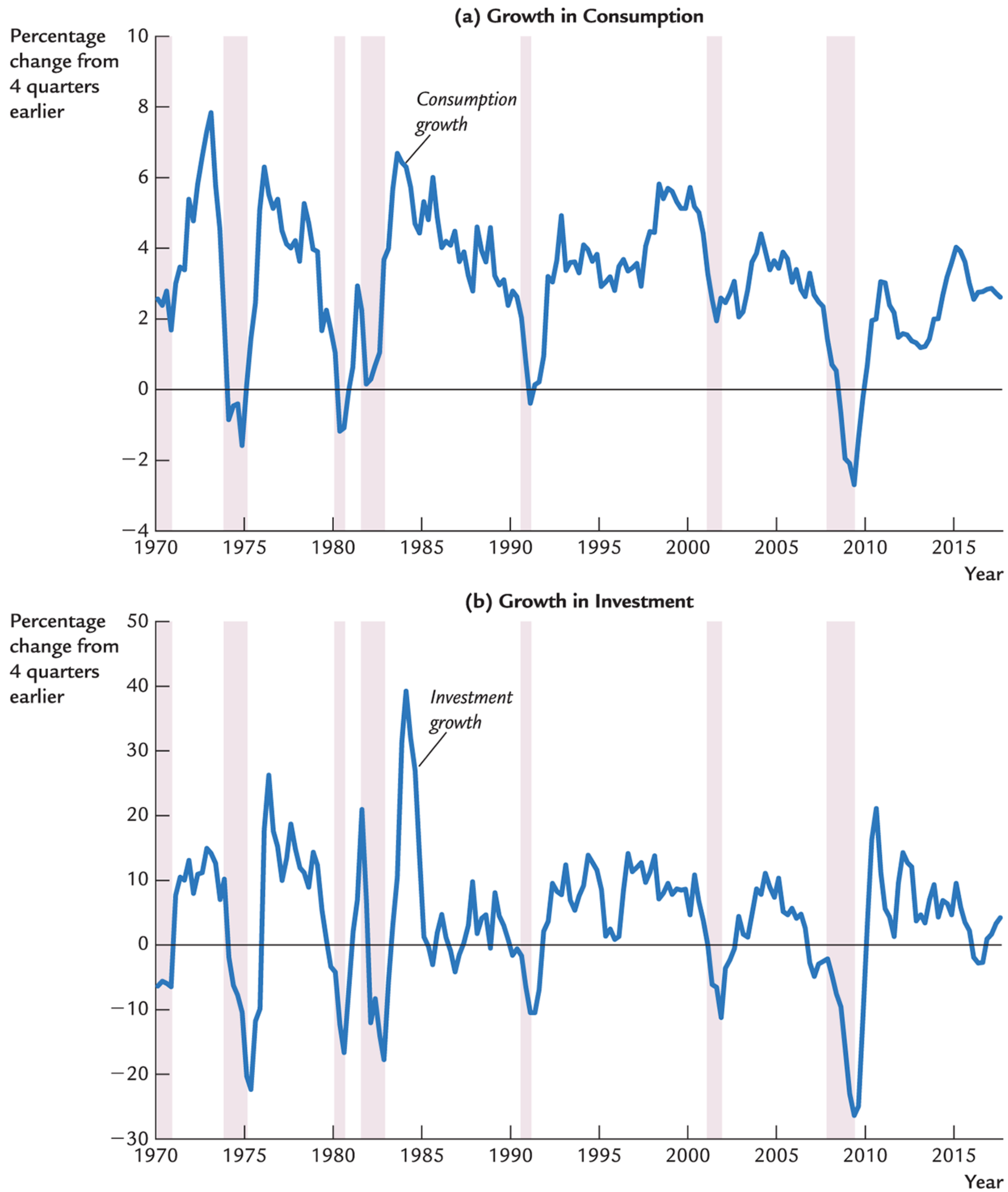


FIGURE 10-2 Growth in Consumption and Investment When the economy heads into a recession, growth in real consumption and investment spending both decline. Investment spending, shown in panel (b), is considerably more volatile than consumption spending, shown in panel (a). The shaded areas represent periods of recession.

Data from: U.S. Department of Commerce, National Bureau of Economic Research.

Unemployment and Okun’s Law

The business cycle is apparent not only in data from the national income accounts but also in data that describe conditions in the labor market. Figure 10-3 shows the unemployment rate from 1970 to 2017, with the shaded areas representing recessions. You can see that unemployment rises in each recession. Other labor-market measures tell a similar story. For example, job vacancies, as measured by the number of help-wanted ads that companies have posted, decline during recessions. Put simply, during economic downturns, jobs are harder to find.

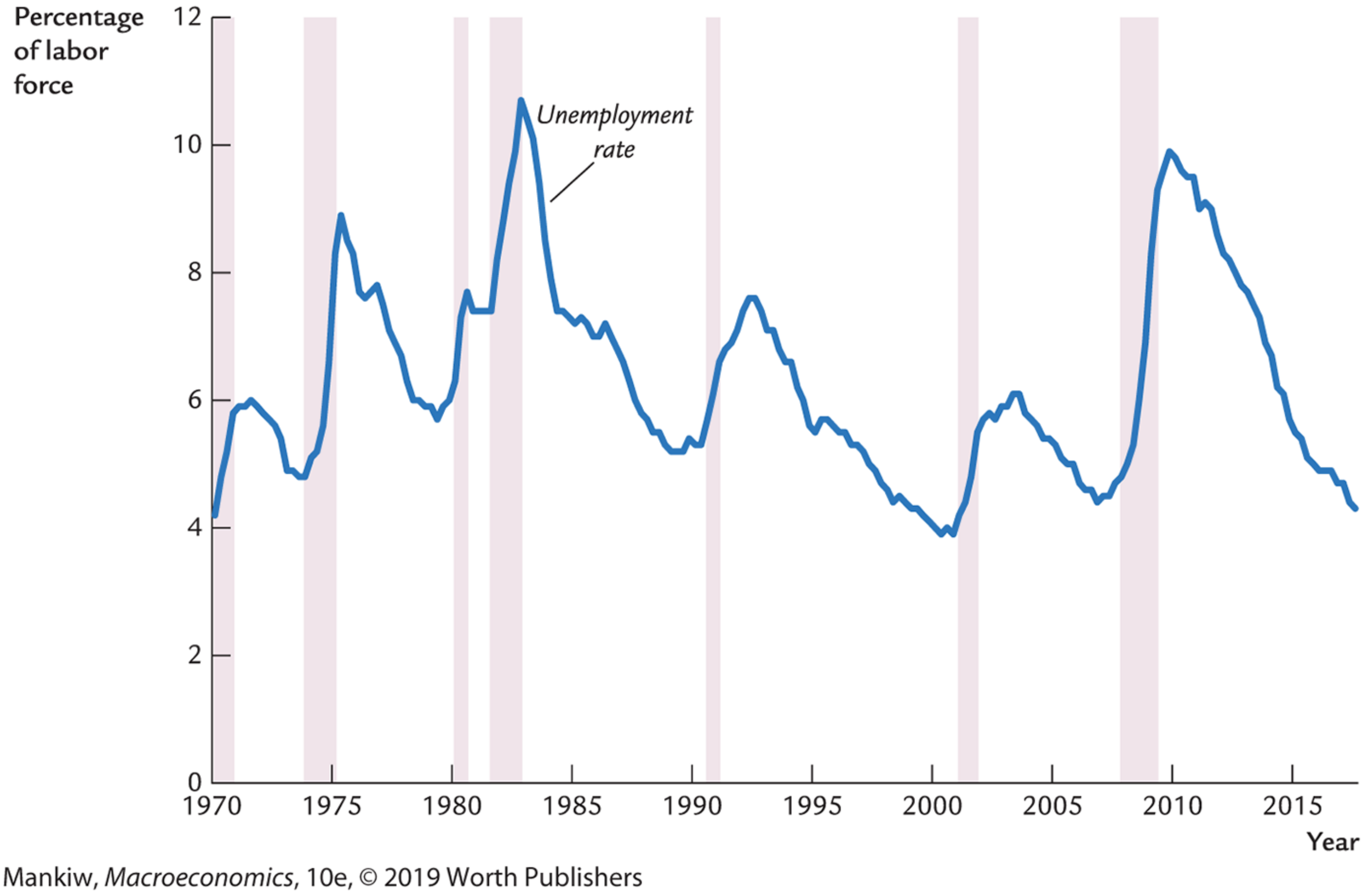
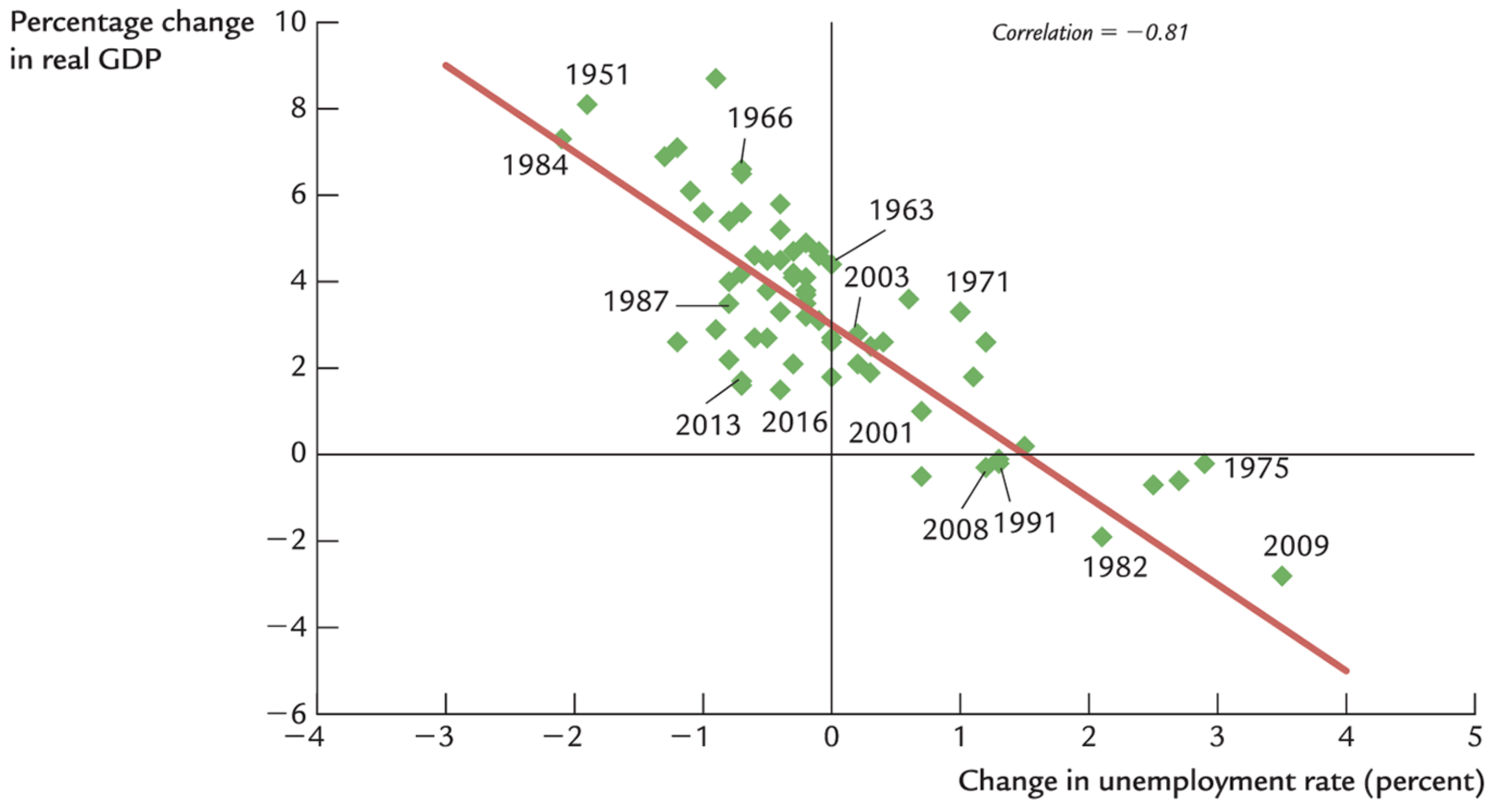


FIGURE 10-3 Unemployment The unemployment rate rises significantly during periods of recession, shown here by the shaded areas.

Data from: U.S. Department of Labor, National Bureau of Economic Research.

What relationship should we expect to find between unemployment and real GDP? Because employed workers help to produce goods and services and unemployed workers do not, increases in the unemployment rate should be associated with decreases in real GDP. This negative relationship between unemployment and GDP is called **Okun’s law**, after Arthur Okun, the economist who first studied it.[2]

Figure 10-4 uses annual data for the United States to illustrate Okun’s law. In this scatterplot, each point represents the data for one year. The horizontal axis represents the change in the unemployment rate from the previous year, and the vertical axis represents the percentage change in GDP. This figure shows clearly that year-to-year changes in the unemployment rate are closely associated with year-to-year changes in real GDP.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 10-4 Okun's Law This figure is a scatterplot of the change in the unemployment rate on the horizontal axis and the percentage change in real GDP on the vertical axis, using data on the U.S. economy. Each point represents one year. The figure shows that increases in unemployment tend to be associated with lower-than-normal growth in real GDP.

Data from: U.S. Department of Commerce, U.S. Department of Labor.

We can be more precise about the magnitude of the Okun's law relationship. The line drawn through the scatter of points tells us that

Percentage Change in Real GDP=3%−2×Change in Unemployment Rate.

$$\text{Percentage Change in Real GDP} = 3\% - 2 \times \text{Change in Unemployment Rate.}$$

If the unemployment rate does not change, real GDP grows by about 3 percent; this normal growth in the production of goods and services is due to growth in the labor force, capital accumulation, and technological progress. In addition, for every percentage point the unemployment rate rises, real GDP growth typically falls by 2 percent. Hence, if the unemployment rate rises from 5 to 7 percent, then real GDP growth would be

Percentage Change in Real GDP=3%−2×(7%−5%)=−1%.

$$\begin{aligned}\text{Percentage Change in Real GDP} &= 3\% - 2 \times (7\% - 5\%) \\ &= -1\%.\end{aligned}$$

In this case, Okun's law says that GDP would fall by 1 percent, indicating that the economy is in a recession.

Okun's law is a reminder that the forces that govern the short-run business cycle are different from those that shape long-run economic growth. As we saw in Chapters 8 and 9, long-run growth in GDP is determined primarily by technological progress. The long-run trend leading to higher standards of living from generation to generation is not associated with any long-run trend in the rate of unemployment. By contrast, short-run movements in GDP are strongly correlated with the utilization of the economy's labor force. The declines in the production of goods and services that occur during recessions are always associated with increases in joblessness.

Leading Economic Indicators

Many economists, particularly those in business and government, are engaged in the task of forecasting short-run fluctuations in the economy. Business economists are interested in forecasting to help their companies plan for changes in the economic environment. Government economists are interested in forecasting for two reasons. First, the economic environment affects the government; for example, the state of the economy influences how much tax revenue the government collects. Second, the government can use monetary and fiscal policy to affect economic activity. Economic forecasts are, therefore, an input into policy planning.

One way that economists arrive at their forecasts is by looking at **leading indicators**, which are variables that often fluctuate before we see movements in the overall economy. Forecasts can differ because economists hold varying opinions about which leading indicators are most reliable.

Each month the Conference Board, a private economics research group, announces the *index of leading economic indicators*. This index includes ten data series that are often used to forecast changes in economic activity about six to nine months into the future. Here is a list of the series:

- *Average weekly hours in manufacturing.* Because businesses often adjust the work hours of existing employees before making new hires or laying off workers, average weekly hours is a leading indicator of employment changes. A longer workweek indicates that firms are asking their employees to work long hours because they are experiencing strong demand for their products; thus, it indicates that firms are likely to increase hiring and production in the future. A shorter workweek indicates weak demand, suggesting that firms are more likely to lay off workers and cut back production.
- *Average weekly initial claims for unemployment insurance.* The number of people making new claims on the unemployment-insurance system is one of the most quickly available indicators of conditions in the labor market. This series is inverted in computing the index of leading indicators so that an increase in the series lowers the index. An increase in the number of people making new claims for unemployment insurance indicates that firms are laying off workers and cutting back production; these layoffs and cutbacks will soon show up in data on employment and production.

- *Manufacturers' new orders for consumer goods and materials.* This indicator is a direct measure of the demand for consumer goods that firms are experiencing. Because an increase in orders depletes a firm's inventories, this statistic typically predicts subsequent increases in production and employment.
- *Manufacturers' new orders for nondefense capital goods, excluding aircraft.* This series is the counterpart to the previous one, but for investment goods rather than consumer goods. When firms experience increased orders, they ramp up production and employment. Aircraft orders are excluded because they are often placed so far in advance of production that these orders contain little information about near-term economic activity.
- *ISM new orders index.* This index, which comes from the Institute for Supply Management, is a third indicator of new orders. It is based on the number of companies reporting increased orders minus the number reporting decreased orders. Unlike the previous two indicators, this one measures the proportion of companies that report rising orders and thus shows whether a change is broadly based. When many firms experience increased orders, higher production and employment will likely soon follow.
- *Building permits for new private housing units.* Construction of new buildings is part of investment—a particularly volatile component of GDP. An increase in building permits means that planned construction is increasing, which indicates a rise in overall economic activity.
- *Index of stock prices.* The stock market reflects expectations about future economic conditions because stock market investors bid up prices when they expect companies to be profitable. An increase in stock prices indicates that investors expect the economy to grow rapidly; a decrease in stock prices indicates that investors expect an economic slowdown.
- *Leading Credit Index.* This component is itself a composite of six financial indicators, such as investor sentiment (based on a survey of stock-market investors) and lending conditions (based on a survey of bank loan officers). When credit conditions are adverse, consumers and businesses find it harder to get the financing they need to make purchases. Thus, a deterioration of credit conditions predicts a decline in spending, production, and employment. This index was added to the leading indicators only recently. The financial crisis of 2008–2009 and subsequent deep recession highlighted the importance of credit conditions for economic activity.
- *Interest rate spread: the yield on 10-year Treasury bonds minus the federal funds rate.* This spread, sometimes called the slope of the yield curve, reflects the market's expectation about future interest rates, which in turn reflect the condition of the economy. A large spread means that interest rates are expected to rise, which typically occurs when economic activity increases.
- *Average consumer expectations for business and economic conditions.* This is a direct measure of expectations, based on two different surveys of households (one conducted by the University of Michigan and one conducted by the Conference Board). Greater optimism about future economic conditions among consumers increases consumer demand for goods and services, suggesting that businesses will expand production and employment to meet the demand.

The index of leading indicators is far from a precise forecast of the future, as short-run economic fluctuations are largely unpredictable. Nonetheless, the index is a useful input into planning by both businesses and the government.

10-2 Time Horizons in Macroeconomics

Having learned some facts that describe short-run economic fluctuations, we turn to our main task in this part of the book: building a theory to explain these fluctuations. That job, it turns out, is not a simple one. It will take us not only the rest of this chapter but also the next four chapters to develop the model of short-run fluctuations in its entirety.

Before we start building the model, however, let's step back and consider a basic question: Why do economists need different models for different time horizons? Why can't we stop the course here and be content with the classical models developed in Chapters 3 through 9? The answer, as we have often noted, is that classical macroeconomic theory applies to the long run but not to the short run. But why is this so?

How the Short Run and the Long Run Differ

Most macroeconomists believe that the key difference between the short run and the long run is the behavior of prices. *In the long run, prices are flexible and can respond to changes in supply or demand. In the short run, many prices are "sticky" at some predetermined level.* Because prices behave differently in the short run than in the long run, various economic events and policies have different effects over different time horizons.

To see how the short run and the long run differ, consider the effects of a change in monetary policy. Suppose that the Federal Reserve suddenly reduces the money supply by 5 percent. According to the classical model, the money supply affects nominal variables—variables measured in terms of money—but not real variables. As you may recall from Chapter 5, the theoretical separation of real and nominal variables is called the *classical dichotomy*, and the irrelevance of the money supply for the determination of real variables is called *monetary neutrality*. Most economists believe that these classical ideas describe how the economy works in the long run: a 5 percent reduction in the money supply lowers all prices (including nominal wages) by 5 percent, while output, employment, and other real variables remain the same. Thus, in the long run, changes in the money supply do not cause fluctuations in output and employment.

In the short run, however, many prices do not respond to changes in monetary policy. A reduction in the money supply does not immediately cause all firms to cut the wages they pay, all stores to change the price tags on their goods, and all restaurants to print new menus. Instead, there is little immediate change in many prices; that is, many prices are sticky. This short-run price stickiness implies that the short-run impact of a change in the money supply is not the same as the long-run impact.

A model of economic fluctuations must take into account this short-run price stickiness. We will see that the failure of prices to adjust quickly and completely to changes in the money supply (as well as to other exogenous changes in economic conditions) means that, in the short run, real variables such as output and employment must do some of the adjusting instead. In other words, during the time horizon over which prices are sticky, the classical dichotomy no longer holds: nominal variables can influence real variables, and the economy can deviate from the equilibrium predicted by the classical model.

CASE STUDY

If You Want to Know Why Firms Have Sticky Prices, Ask Them

How sticky are prices, and why are they sticky? In an intriguing study, economist Alan Blinder tackled these questions directly by surveying firms about their price-adjustment decisions.

Blinder began by asking firm managers how often they changed prices. The answers, summarized in Table 10-1, yielded two conclusions. First, sticky prices are common. The typical firm in the economy adjusts its prices once or twice a year. Second, there are large differences among firms in the frequency of price adjustment. About 10 percent of firms changed prices more than once a week, and about the same number changed prices less than once a year.

TABLE 10-1 The Frequency of Price Adjustment

This table is based on answers to the question: How often do the prices of your most important products change in a typical year?

Frequency	Percentage of Firms
Less than once	10.2
Once	39.3
1.01 to 2	15.6
2.01 to 4	12.9
4.01 to 12	7.5
12.01 to 52	4.3
52.01 to 365	8.6
More than 365	1.6

Data from: Alan S. Blinder, “On Sticky Prices: Academic Theories Meet the Real World,” in N. G. Mankiw, ed., *Monetary Policy* (Chicago: University of Chicago Press, 1994), 117–154, Table 4.1.

Blinder then asked the firm managers why they didn’t change prices more often. In particular, he explained to the managers several economic theories of sticky prices and asked them to judge how well each of these theories described their firms. Table 10-2 summarizes the theories and ranks them by the percentage of managers who accepted each theory as an accurate description of their firms’ pricing decisions. Notice that each of the theories was endorsed by some of the managers, but each was rejected by a large number as well. Perhaps different theories apply to different firms, depending on industry characteristics. Price stickiness may be a macroeconomic phenomenon without a single microeconomic explanation.

TABLE 10-2 Theories of Price Stickiness

Theory and Brief Description	Percentage of Managers Who Accepted Theory
Coordination failure: Firms hold back on price changes, waiting for others to go first	60.6
Cost-based pricing with lags: Price increases are delayed until costs rise	55.5
Delivery lags, service, etc.: Firms prefer to vary other product attributes, such as delivery lags, service, and product quality	54.8
Implicit contracts: Firms tacitly agree to stabilize prices, perhaps out of “fairness” to customers	50.5
Nominal contracts: Prices are fixed by explicit contracts	35.7
Costs of price adjustment: Firms incur costs of changing prices	30.0
Procyclical elasticity: Demand curves become less elastic as they shift in	29.7
Pricing points: Certain prices (like $9.99) have special psychological significance	24.0
Inventories: Firms vary inventory stocks instead of prices	20.9
Constant marginal cost: Marginal cost is flat and markups are constant	19.7
Hierarchical delays:	13.6

Bureaucratic delays slow down decisions	
Judging quality by price: Firms fear customers will mistake price cuts for reductions in quality	10.0

Data from: Alan S. Blinder, "On Sticky Prices: Academic Theories Meet the Real World," in N. G. Mankiw, ed., *Monetary Policy* (Chicago: University of Chicago Press, 1994), 117–154, Tables 4.3 and 4.4.

Among the dozen theories, coordination failure tops the list. According to Blinder, this is an important finding because it suggests that the inability of firms to coordinate price changes plays a key role in explaining price stickiness and, thus, short-run economic fluctuations. He writes, "The most obvious policy implication of the model is that more coordinated wage and price setting—somehow achieved—could improve welfare. But if this proves difficult or impossible, the door is opened to activist monetary policy to cure recessions."[3] ■

The Model of Aggregate Supply and Aggregate Demand

How does the introduction of sticky prices change our view of how the economy works? We can answer this question by considering economists' two favorite words: supply and demand.

In classical macroeconomic theory, the economy's output depends on its ability to *supply* goods and services, which in turn depends on the supplies of capital and labor and on the available production technology. This is the essence of the basic classical model in Chapter 3, as well as of the Solow growth model in Chapters 8 and 9. Flexible prices are a crucial assumption of classical theory. The theory posits, sometimes implicitly, that prices adjust to ensure that the quantity of output demanded equals the quantity supplied.

The economy works quite differently when prices are sticky. In this case, as we will see, output also depends on the economy's *demand* for goods and services. Demand, in turn, depends on many factors: consumers' confidence about their economic prospects, firms' perceptions about the profitability of new investments, and monetary and fiscal policy. Because monetary and fiscal policy can influence demand, and demand can influence the economy's output over the time horizon when prices are sticky, price stickiness provides a rationale for why these policies may be useful in stabilizing the economy in the short run.

In the rest of this chapter, we begin developing a model that makes these ideas more precise. The place to start is the model of supply and demand, which we used in Chapter 1 to discuss the market for pizza. This basic model offers some of the most fundamental insights in economics. It shows how the supply and demand for any good jointly determine the good's price and the quantity sold, as well as how shifts in supply and demand affect the price and quantity. We now introduce the "economy-size" version of this model—*the model of aggregate supply and aggregate demand*. This macroeconomic model allows us to study how the aggregate price level and the quantity of aggregate output are determined in the short run. It also provides a way to contrast how the economy behaves in the long run and how it behaves in the short run.

Although the model of aggregate supply and aggregate demand resembles the model of supply and demand for a single good, the analogy is not exact. The model of supply and demand for a single good considers only one good within a large economy. By contrast, as we will see in the coming chapters, the model of aggregate supply and aggregate demand is a sophisticated model that incorporates the interactions among many markets. In the remainder of this chapter, we get a first glimpse at those interactions by examining the model in its simplest form. Our goal here is not to explain the model fully but to introduce its key elements and show how it can help explain short-run fluctuations.

10-3 Aggregate Demand

Aggregate demand (*AD*) is the relationship between the quantity of output demanded and the aggregate price level. In other words, the aggregate demand curve tells us the quantity of goods and services people want to buy at any given level of prices. We examine the theory of aggregate demand in detail in Chapters 11 through 13. Here we use the quantity theory of money to provide a simple, although incomplete, derivation of the aggregate demand curve.

The Quantity Equation as Aggregate Demand

Recall from Chapter 5 that the quantity theory says that

$$MV = PY,$$

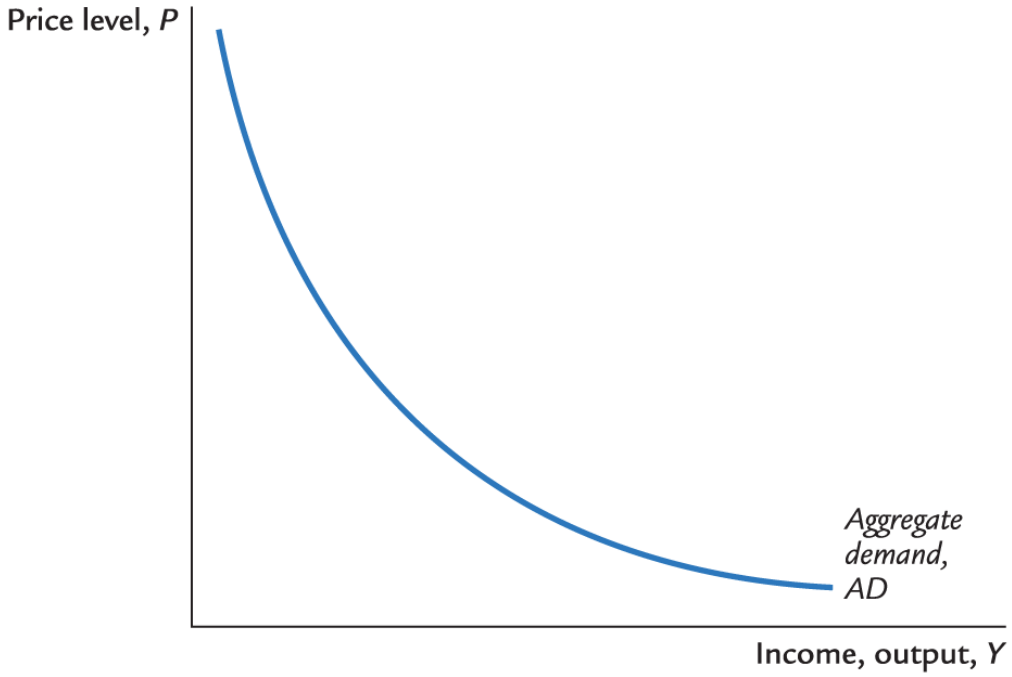
where *M* is the money supply, *V* is the velocity of money, *P* is the price level, and *Y* is the amount of output. If the velocity of money is constant, then this equation states that the money supply determines the nominal value of output, which in turn is the product of the price level and the amount of output.

When interpreting this equation, it is useful to recall that the quantity equation can be rewritten in terms of the supply and demand for real money balances:

$$M/P = (M/P)^d = kY,$$

where $k = 1/V$ is a parameter representing how much money people want to hold for every dollar of income. In this form, the quantity equation states that the supply of real money balances M/P equals the demand for real money balances $(M/P)^d$ and that the demand is proportional to output *Y*. The velocity of money *V* is the flip side of the money demand parameter *k*. The assumption of constant velocity is equivalent to the assumption of a constant demand for real money balances per unit of output.

If we assume that velocity *V* is constant and the money supply *M* is fixed by the central bank, then the quantity equation yields a negative relationship between the price level *P* and output *Y*. Figure 10-5 graphs the combinations of *P* and *Y* that satisfy the quantity equation holding *M* and *V* constants. This downward-sloping curve is called the aggregate demand curve.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 10-5 The Aggregate Demand Curve The aggregate demand curve *AD* shows the relationship between the price level *P* and the quantity of goods and services demanded *Y*. It is drawn for a given value of the money supply *M*. The aggregate demand curve slopes downward: the higher the price level *P*, the lower the level of real balances M/P, M/P, and therefore the lower the quantity of goods and services demanded *Y*.

Why the Aggregate Demand Curve Slopes Downward

As a strictly mathematical matter, the quantity equation explains the downward slope of the aggregate demand curve very simply. The money supply *M* and the velocity of money *V* determine the nominal value of output *PY*. Once *PY* is fixed, if *P* goes up, *Y* must go down.

What is the economic intuition that lies behind this mathematical relationship? For a complete explanation of the downward slope of the aggregate demand curve, we have to wait for a couple of chapters. For now, however, consider the following logic: because we have assumed the velocity of money is fixed, the money supply determines the dollar value of all transactions in the economy. (This conclusion should be familiar from Chapter 5.) If the price level rises, each transaction requires more dollars, so the number of transactions and thus the quantity of goods and services purchased must fall.

We can also explain the downward slope of the aggregate demand curve by thinking about the supply and demand for real money balances. If output is higher, people engage in more transactions and need higher real balances M/P. M/P. For a fixed money supply *M*, higher real balances imply a lower price level. Conversely, if the price level is lower, real money balances are higher; the higher level of real balances allows a greater volume of transactions, which means a greater quantity of output is demanded.

Shifts in the Aggregate Demand Curve

The aggregate demand curve is drawn for a fixed money supply. In other words, it tells us the possible combinations of P and Y for a given value of M. If the Fed changes the money supply, then the possible combinations of P and Y change, which means the aggregate demand curve shifts.

For example, consider what happens if the Fed reduces the money supply. The quantity equation, $MV = PY$, tells us that the reduction in the money supply leads to a proportionate reduction in the nominal value of output PY. For any given price level, the amount of output is lower, and for any given amount of output, the price level is lower. As in Figure 10-6 panel (a), the aggregate demand curve relating P and Y shifts inward.

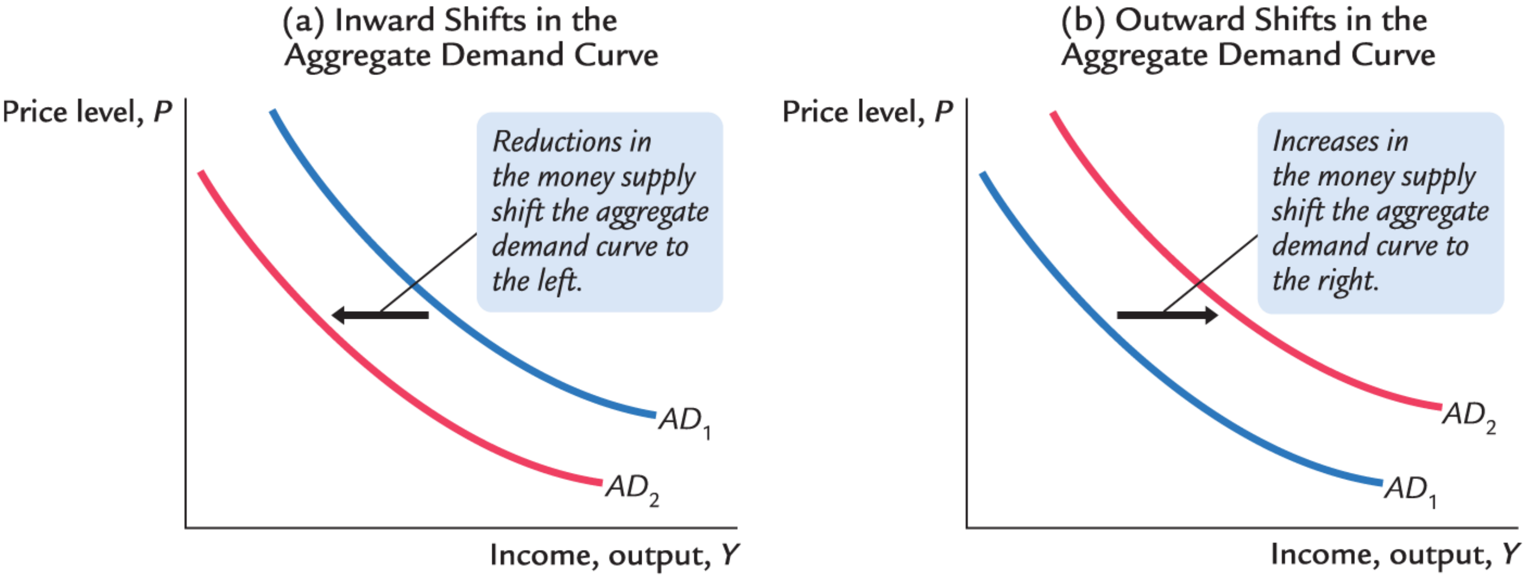


FIGURE 10-6 Shifts in the Aggregate Demand Curve Changes in the money supply shift the aggregate demand curve. In panel (a), a decrease in the money supply M reduces the nominal value of output PY. For any given price level P, output Y is lower. Thus, a decrease in the money supply shifts the aggregate demand curve inward, from AD_1 to AD_2. In panel (b), an increase in the money supply M raises the nominal value of output PY. For any given price level P, output Y is higher. Thus, an increase in the money supply shifts the aggregate demand curve outward, from AD_1 to AD_2.

The opposite occurs if the Fed increases the money supply. The quantity equation tells us that an increase in M leads to an increase in PY. For any given price level, the amount of output is higher, and for any given amount of output, the price level is higher. As shown in Figure 10-6 panel (b), the aggregate demand curve shifts outward.

Although the quantity theory of money provides a simple explanation of the aggregate demand curve, be forewarned that reality is more complex. Fluctuations in the money supply are not the only source of fluctuations in aggregate demand. Even if the money supply is held constant, the aggregate demand curve shifts if some event causes a change in the velocity of money. Over the next two chapters, we develop a more general model of aggregate demand, called the *IS–LM model*, which will allow us to consider many possible reasons for shifts in the aggregate demand curve.

10-4 Aggregate Supply

By itself, the aggregate demand curve does not tell us the price level or the amount of output that will prevail in the economy; it merely shows a relationship between these two variables. To accompany the aggregate demand curve, we need another relationship between *P* and *Y* that crosses the aggregate demand curve—an aggregate supply curve. The aggregate demand and aggregate supply curves together pin down the economy's price level and quantity of output.

Aggregate supply (*AS*) is the relationship between the quantity of goods and services supplied and the price level. Because the firms that supply goods and services have flexible prices in the long run but sticky prices in the short run, the aggregate supply relationship depends on the time horizon. We need to introduce two different aggregate supply curves: the long-run aggregate supply curve *LRAS* and the short-run aggregate supply curve *SRAS*. We also need to discuss how the economy makes the transition from the short run to the long run.

The Long Run: The Vertical Aggregate Supply Curve

Because the classical model describes how the economy behaves in the long run, we derive the long-run aggregate supply curve from the classical model. Recall from Chapter 3 that the amount of output produced depends on the fixed amounts of capital and labor and on the available technology. To show this, we write

Y=F(K‾,L‾) =Y‾.

$$\begin{aligned} Y &= F(\bar{K}, \bar{L}) \\ &= \bar{Y}. \end{aligned}$$

According to the classical model, output does not depend on the price level. To show that output is fixed at this level, regardless of the price level, we draw a vertical aggregate supply curve, as in Figure 10-7. In the long run, the intersection of the aggregate demand curve with this vertical aggregate supply curve determines the price level.

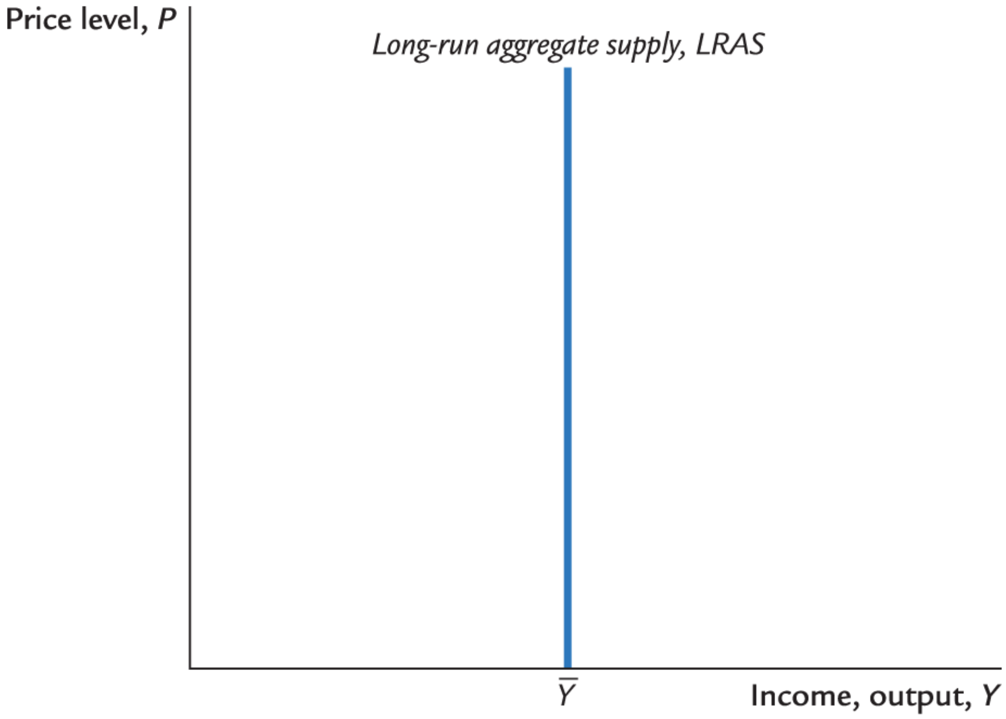


FIGURE 10-7 The Long-Run Aggregate Supply Curve In the long run, output is determined by the amounts of capital and labor and by the available technology; it does not depend on the price level. Therefore, the long-run aggregate supply curve, *LRAS*, is vertical.

If the aggregate supply curve is vertical, then changes in aggregate demand affect prices but not output. For example, if the money supply falls, the aggregate demand curve shifts downward, as in Figure 10-8. The economy moves from the old intersection of aggregate supply and aggregate demand, point A, to the new intersection, point B. The shift in aggregate demand affects only prices.

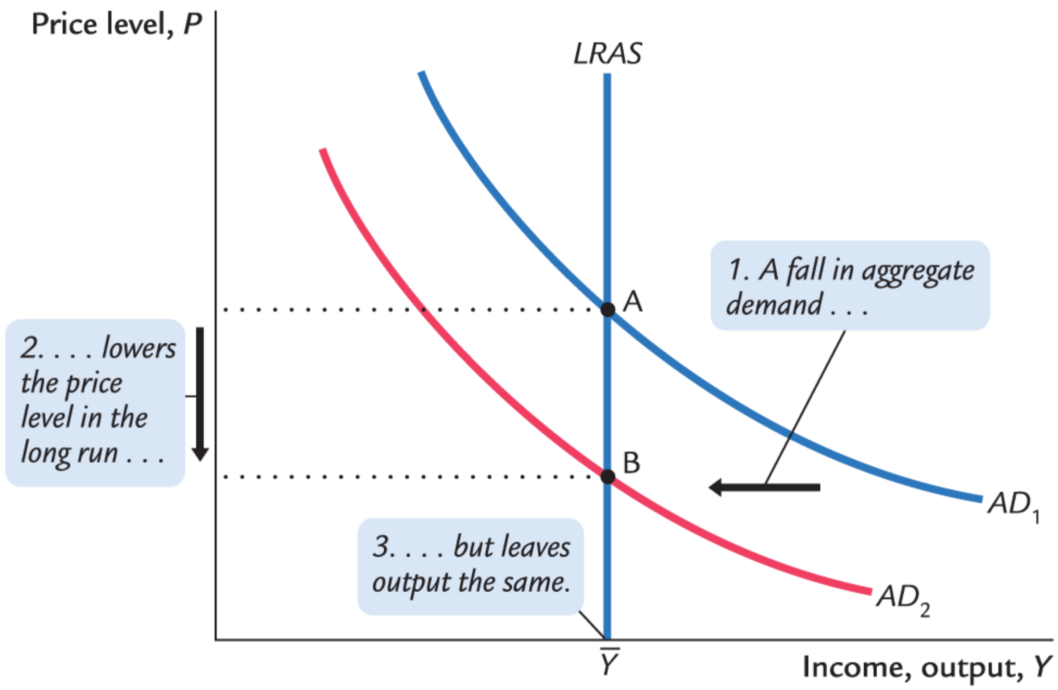


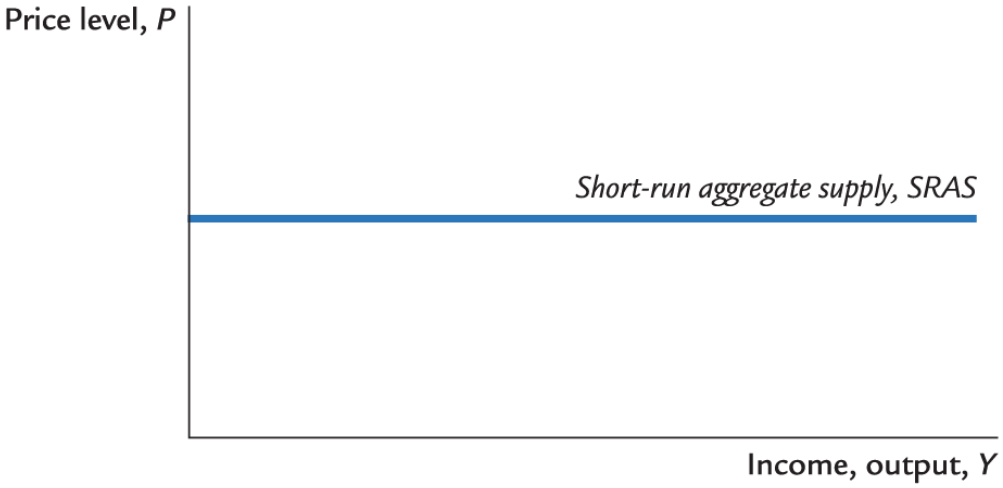
FIGURE 10-8 Shifts in Aggregate Demand in the Long Run A reduction in the money supply shifts the aggregate demand curve downward, from AD_1 to AD_2. The equilibrium for the economy moves from point A to point B. Because the aggregate supply curve is vertical in the long run, the reduction in aggregate demand affects the price level but not output.

The vertical aggregate supply curve satisfies the classical dichotomy because it implies that the money supply does not affect output. The long-run level of output $\mathrm{Y}^{-}\overline{Y}$ is called the *full-employment*, or *natural*, level of output. It is the level of output at which the economy's resources are fully employed or, more realistically, at which unemployment is at its natural rate.

The Short Run: The Horizontal Aggregate Supply Curve

The classical model and the vertical aggregate supply curve apply only in the long run. In the short run, some prices are sticky and therefore do not adjust to changes in demand. Because of this price stickiness, the short-run aggregate supply curve is not vertical.

In this chapter, we simplify things by assuming an extreme example. Suppose that all firms have issued price catalogs and it is too costly for them to issue new ones. Thus, all prices are stuck at predetermined levels. At these prices, firms are willing to sell as much as their customers are willing to buy, and they hire just enough labor to produce the amount demanded. Because the price level is fixed, we represent this situation in Figure 10-9 with a horizontal aggregate supply curve.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 10-9 The Short-Run Aggregate Supply Curve In this extreme example, all prices are fixed in the short run. Therefore, the short-run aggregate supply curve, *SRAS*, is horizontal.

The short-run equilibrium of the economy is the intersection of the aggregate demand curve and this horizontal short-run aggregate supply curve. In this case, changes in aggregate demand affect output. For example, if the Fed suddenly reduces the money supply, the aggregate demand curve shifts inward, as in Figure 10-10. The economy moves from the old intersection of aggregate demand and aggregate supply, point A, to the new intersection, point B. The movement from point A to point B represents a decline in output at a fixed price level.

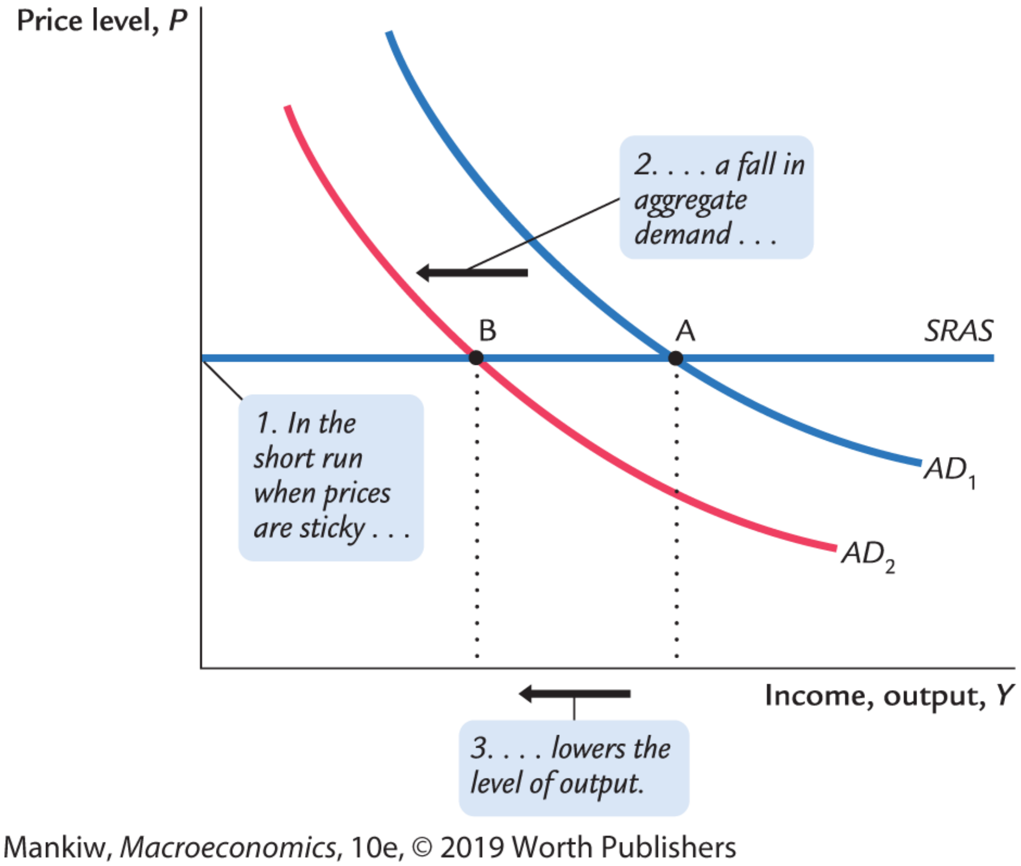


FIGURE 10-10 Shifts in Aggregate Demand in the Short Run A reduction in the money supply shifts the aggregate demand curve downward, from AD_1 to AD_2. The equilibrium for the economy moves from point A to point B. Because the aggregate supply curve is horizontal in the short run, the reduction in aggregate demand reduces output.

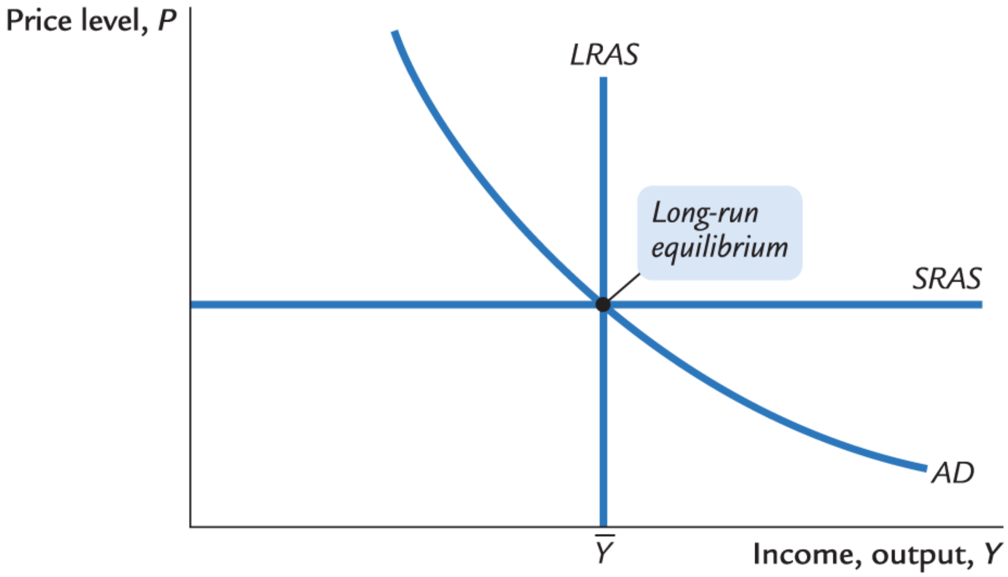
Thus, a fall in aggregate demand reduces output in the short run because prices do not adjust instantly. After the sudden fall in aggregate demand, firms are stuck with prices that are too high. With demand low and prices high, firms sell less of their product, so they reduce production and lay off workers. The economy experiences a recession.

Once again, be forewarned that reality is a bit more complicated than illustrated here. Although many prices are sticky in the short run, other prices can respond quickly to changing circumstances. As we will see in Chapter 14, in an economy with some sticky prices and some flexible prices, the short-run aggregate supply curve is upward sloping rather than horizontal. Figure 10-10 shows the extreme case in which all prices are stuck. Because this case is simpler, it is a useful starting point for thinking about short-run aggregate supply.

From the Short Run to the Long Run

We can summarize our analysis so far as follows: *over long periods of time, prices are flexible, the aggregate supply curve is vertical, and changes in aggregate demand affect the price level but not output. Over short periods of time, prices are sticky, the aggregate supply curve is flat, and changes in aggregate demand do affect the economy's output of goods and services*.

How does the economy make the transition from the short run to the long run? Let's trace the effects of a fall in aggregate demand over time. Suppose that the economy begins in long-run equilibrium, as shown in Figure 10-11. In this figure, there are three curves: the aggregate demand curve, the long-run aggregate supply curve, and the short-run aggregate supply curve. The long-run equilibrium is the point at which aggregate demand crosses the long-run aggregate supply curve. Prices have adjusted to reach this equilibrium. Therefore, when the economy is in its long-run equilibrium, the short-run aggregate supply curve must cross this point as well.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 10-11 Long-Run Equilibrium In the long run, the economy finds itself at the intersection of the long-run aggregate supply curve and the aggregate demand curve. Because prices have adjusted to reach this equilibrium, the short-run aggregate supply curve crosses this point as well.

Now suppose that the Fed reduces the money supply and the aggregate demand curve shifts downward, as in Figure 10-12. In the short run, prices are sticky, so the economy moves from point A to point B. Output and employment fall below their natural levels, which means the economy is in a recession. Over time, in response to the low demand, wages and prices fall. The gradual reduction in the price level moves the economy downward along the aggregate demand curve to point C, which is the new long-run equilibrium. In this new long-run equilibrium (point C), output and employment are back to their natural levels, but prices are lower than in the old long-run equilibrium (point A). Thus, a shift in aggregate demand affects output in the short run, but this effect dissipates over time as firms adjust their prices.

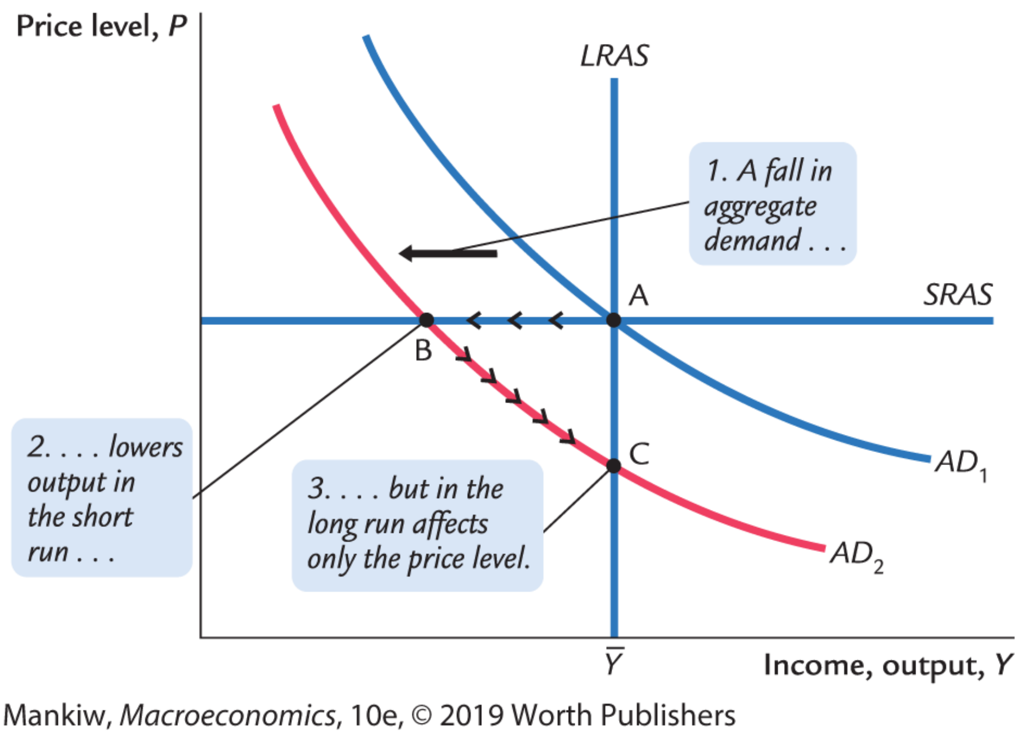


FIGURE 10-12 A Reduction in Aggregate Demand The economy begins in long-run equilibrium at point A. A reduction in aggregate demand, perhaps caused by a decrease in the money supply, moves the economy from point A to point B, where output is below its natural level. As prices fall, the economy gradually recovers from the recession, moving from point B to point C.

CASE STUDY

A Monetary Lesson from French History

Finding modern examples to illustrate the lessons from Figure 10-12 is hard. Modern central banks are too smart to engineer a substantial reduction in the money supply for no good reason. They know that a recession would ensue, and they usually do their best to prevent that from happening. Fortunately, history often fills in the gap when recent experience fails to produce the right experiment.

A vivid example of the effects of monetary contraction occurred in eighteenth-century France. In 2009, François Velde, an economist at the Federal Reserve Bank of Chicago, studied this episode in French economic history.

The story begins with the unusual nature of French money at the time. The money stock in this economy included a variety of gold and silver coins that, in contrast to modern money, did not indicate specific monetary values. Instead, the monetary value of each coin was set by government decree, and the government could easily change the monetary value and thus the money supply. Sometimes this would occur literally overnight. It is almost as if, while you were sleeping, every $1 bill in your wallet was replaced by a bill worth only 80 cents.

Indeed, that is what happened on September 22, 1724. Every person in France woke up with 20 percent less money than he had had the night before. Over the course of seven months, the nominal value of the money stock was reduced by about 45 percent. The goal of these changes was to reduce prices in the economy to what the government considered an appropriate level.

What happened as a result of this policy? Velde reports the following consequences:

> Although prices and wages did fall, they did not do so by the full 45 percent; moreover, it took them

> months, if not years, to fall that far. Real wages in fact rose, at least initially. Interest rates rose. The only market that adjusted instantaneously and fully was the foreign exchange market. Even markets that were as close to fully competitive as one can imagine, such as grain markets, failed to react initially. . . .
>
> At the same time, the industrial sector of the economy (or at any rate the textile industry) went into a severe contraction, by about 30 percent. The onset of the recession may have occurred before the deflationary policy began, but it was widely believed at the time that the severity of the contraction was due to monetary policy, in particular to a resulting "credit crunch" as holders of money stopped providing credit to trade in anticipation of further price declines (the "scarcity of money" frequently blamed by observers). Likewise, it was widely believed (on the basis of past experience) that a policy of inflation would halt the recession, and coincidentally or not, the economy rebounded once the nominal money supply was increased by 20 percent in May 1726.

This description of events from French history fits well with the lessons from mainstream macroeconomic theory.[4]
■

10-5 Stabilization Policy

Fluctuations in the economy as a whole come from changes in aggregate supply or aggregate demand. Economists call exogenous events that shift these curves **shocks** to the economy. A shock that shifts the aggregate demand curve is called a **demand shock**, and a shock that shifts the aggregate supply curve is called a **supply shock**. These shocks disrupt the economy by pushing output and employment away from their natural levels. One goal of the model of aggregate supply and aggregate demand is to show how shocks cause economic fluctuations.

Another goal of the model is to evaluate how macroeconomic policy can respond to these shocks. Economists use the term **stabilization policy** to refer to policy actions aimed at reducing the severity of short-run economic fluctuations. Because output and employment fluctuate around their long-run natural levels, stabilization policy dampens the business cycle by keeping output and employment as close to their natural levels as possible.

In the coming chapters, we examine in detail how stabilization policy works and what practical problems arise in its use. Here we begin our analysis of stabilization policy using our simplified version of the model of aggregate demand and aggregate supply. In particular, we examine how monetary policy might respond to shocks. Monetary policy is an important component of stabilization policy because, as we have seen, the money supply has a powerful impact on aggregate demand.

Shocks to Aggregate Demand

Consider an example of a demand shock: the introduction and expanded availability of credit cards. Because using credit cards is often a more convenient way to make purchases than using cash, credit cards reduce the quantity of money that people choose to hold. This reduction in money demand is equivalent to an increase in the velocity of money. When each person holds less money, the money demand parameter *k* falls. This means that each dollar of money moves from hand to hand more quickly, so velocity V(= 1/k)$V(= 1/k)$ rises.

If the money supply is held constant, the increase in velocity causes nominal spending to rise and the aggregate demand curve to shift outward, as in Figure 10-13. In the short run, the increase in demand raises the output of the economy—it causes an economic boom. At the old prices, firms now sell more output. Therefore, they hire more workers, ask their existing workers to work longer hours, and make greater use of their factories and equipment.

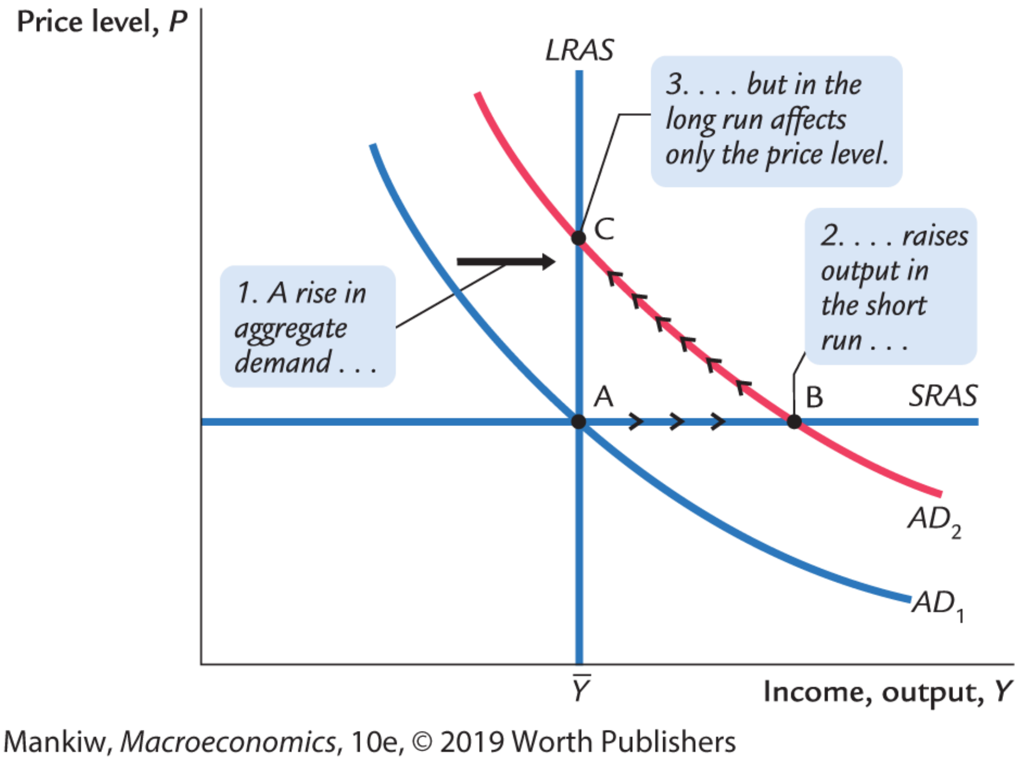


FIGURE 10-13 An Increase in Aggregate Demand The economy begins in long-run equilibrium at point A. An increase in aggregate demand, perhaps due to an increase in the velocity of money, moves the economy from point A to point B, where output is above its natural level. As prices rise, output gradually returns to its natural level, and the economy moves from point B to point C.

Over time, the high level of aggregate demand pulls up wages and prices. As the price level rises, the quantity of output demanded declines, and the economy gradually approaches the natural level of production. But during the transition to the higher price level, output is higher than its natural level.

What can the Fed do to dampen this boom and keep output closer to the natural level? The Fed might reduce the money supply to offset the increase in velocity. Offsetting the change in velocity would stabilize aggregate demand. Thus, the Fed can reduce or even eliminate the impact of demand shocks on output and employment if it can skillfully control the money supply. Whether the Fed in fact has the necessary skill is a more difficult question, which we take up in Chapter 16.

Shocks to Aggregate Supply

Shocks to aggregate supply can also cause economic fluctuations. A supply shock is a shock to the economy that alters the cost of producing goods and services and, as a result, the prices that firms charge. Because supply shocks have a direct impact on the price level, they are sometimes called *price shocks*. Here are some examples:

- A drought that destroys crops. The reduction in food supply pushes up food prices.
- A new environmental protection law that requires firms to reduce their emissions of pollutants. Firms

pass on the added costs to customers in the form of higher prices.

- An increase in union aggressiveness. This pushes up wages and the prices of the goods produced by union workers.
- The organization of an international oil cartel. By curtailing competition, the major oil producers can raise the world price of oil.

These events are *adverse* supply shocks, which means they push costs and prices upward. A *favorable* supply shock, such as the breakup of an international oil cartel, reduces costs and prices.

Figure 10-14 shows how an adverse supply shock affects the economy. The short-run aggregate supply curve shifts upward. (The supply shock may also lower the natural level of output and shift the long-run aggregate supply curve to the left, but we ignore that effect here.) If aggregate demand is held constant, the economy moves from point A to point B: the price level rises, and output falls below its natural level. This experience is called *stagflation* because it combines economic stagnation (falling output and, from Okun's law, rising unemployment) with inflation (rising prices).

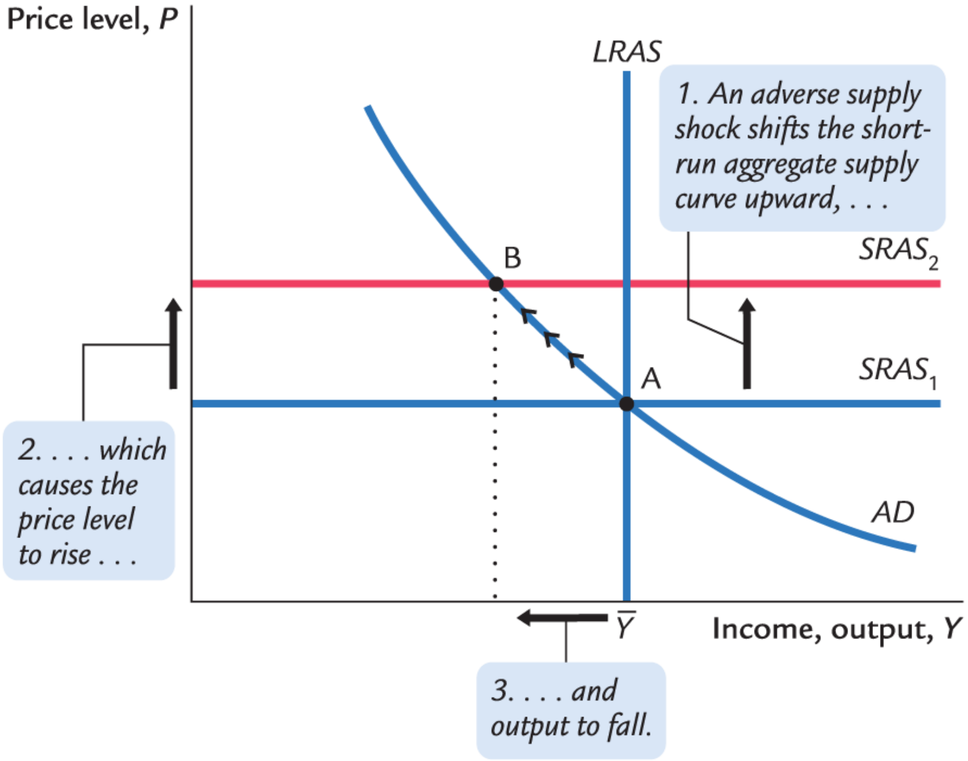


FIGURE 10-14 An Adverse Supply Shock An adverse supply shock pushes up costs and thus prices. If aggregate demand is held constant, the economy moves from point A to point B, leading to stagflation—a combination of increasing prices and falling output. Eventually, as prices fall, the economy returns to the natural level of output, point A.

Faced with an adverse supply shock, a policymaker with the ability to influence aggregate demand, such as the Fed, has a hard choice between two options. The first option, implicit in Figure 10-14, is to hold aggregate demand constant. In this case, output and employment are lower than the natural level. Eventually, prices will fall to restore full employment at the old price level (point A), but the cost of this adjustment process is a painful recession.

The second option, shown in Figure 10-15, is to expand aggregate demand to bring the economy toward the natural level of output more quickly. If the increase in aggregate demand coincides with the shock to aggregate supply, the economy goes immediately from point A to point C. In this case, the Fed is said to *accommodate* the supply shock. The drawback of this option is that the price level is permanently higher. There is no way to adjust aggregate demand to maintain full employment and keep the price level stable.

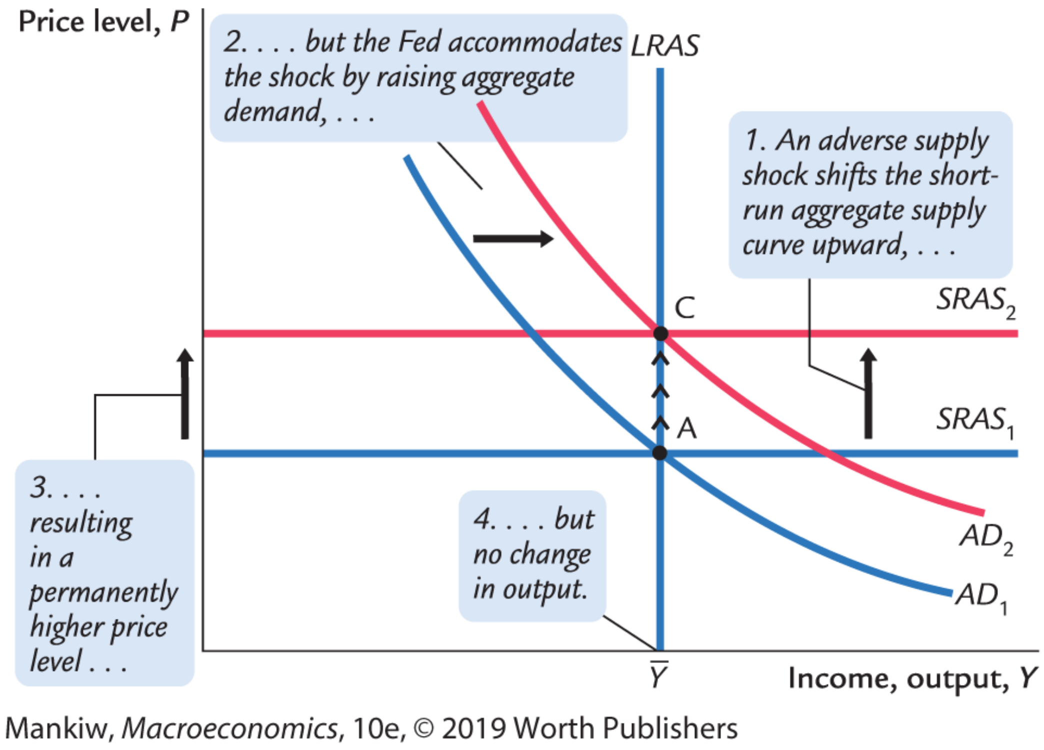


FIGURE 10-15 Accommodating an Adverse Supply Shock In response to an adverse supply shock, the Fed can increase aggregate demand to prevent a reduction in output. The economy moves from point A to point C. The cost of this policy is a permanently higher level of prices.

CASE STUDY

How OPEC Helped Cause Stagflation in the 1970s and Euphoria in the 1980s

The most disruptive supply shocks in recent history were caused by OPEC, the Organization of the Petroleum Exporting Countries. OPEC is a cartel, which is an organization of suppliers that coordinate production levels and prices. In the early 1970s, OPEC's reduction in the supply of oil nearly doubled the world price. This increase in oil prices caused stagflation in most industrial countries. These statistics show what happened in the United States:

Year	Change in Oil Prices	Inflation Rate (CPI)	Unemployment Rate
1973	11.0%	6.2%	4.9%
1974	68.0	11.0	5.6
1975	16.0	9.1	8.5
1976	3.3	5.8	7.7
1977	8.1	6.5	7.1

The 68 percent increase in the price of oil in 1974 was an adverse supply shock of major proportions. As we would expect, this shock led to both higher inflation and higher unemployment.

A few years later, when the world economy had nearly recovered from the first OPEC recession, almost the same thing happened again. OPEC raised oil prices, causing further stagflation. Here are the statistics for the United States:

Year	Change in Oil Prices	Inflation Rate (CPI)	Unemployment Rate
1978	9.4%	7.7%	6.1%
1979	25.4	11.3	5.8
1980	47.8	13.5	7.0
1981	44.4	10.3	7.5
1982	−8.7	6.1	9.5

The increases in oil prices in 1979, 1980, and 1981 again led to double-digit inflation and higher unemployment.

In the mid-1980s, political turmoil among the Arab countries weakened OPEC's ability to restrain supplies of oil. Oil prices fell, reversing the stagflation of the 1970s and the early 1980s. Here's what happened:

Year	Changes in Oil Prices	Inflation Rate (CPI)	Unemployment Rate
1983	−7.1%	3.2%	9.5%
1984	−1.7	4.3	7.4
1985	−7.5	3.6	7.1
1986	−44.5	1.9	6.9
1987	18.3	3.6	6.1

In 1986 oil prices fell by nearly half. This favorable supply shock led to one of the lowest inflation rates experienced during that era and also to falling unemployment.

More recently, OPEC has not been a major cause of economic fluctuations. Conservation efforts and technological changes that improve energy efficiency have made the U.S. economy less susceptible to oil shocks. Moreover, the economy today is based less on manufacturing and more on services, which require less energy to produce. From 1980 to 2016, the amount of oil consumed per unit of real GDP fell by 55 percent. As a result, fluctuations in oil prices now have a smaller impact on the economy.[5] ■

10-6 Conclusion

This chapter has introduced a framework to study economic fluctuations: the model of aggregate supply and aggregate demand. The model is built on the assumption that prices are sticky in the short run and flexible in the long run. It shows how shocks to the economy cause output to deviate temporarily from the level implied by the classical model.

The model also highlights the role of monetary policy. On the one hand, poor monetary policy can be a source of destabilizing shocks to the economy. On the other hand, a well-run monetary policy can respond to shocks and stabilize the economy.

In the chapters that follow, we refine our understanding of this model and our analysis of stabilization policy. Chapters 11 through 13 go beyond the quantity equation to refine our theory of aggregate demand. Chapter 14 examines aggregate supply in more detail. The rest of the book then uses this model as the platform from which to dive into more advanced topics in macroeconomic theory and policy.

CHAPTER 11

Aggregate Demand I: Building the *IS–LM* Model

Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

> I shall argue that the postulates of the classical theory are applicable to a special case only and not to the general case. . . . Moreover, the characteristics of the special case assumed by the classical theory happen not to be those of the economic society in which we actually live, with the result that its teaching is misleading and disastrous if we attempt to apply it to the facts of experience.
>
> —John Maynard Keynes, *The General Theory*

Of all the economic fluctuations in world history, the one that stands out as particularly large, painful, and intellectually significant is the Great Depression of the 1930s. During that time, the United States and many other countries experienced massive unemployment and greatly reduced incomes. In the worst year, 1933, one-fourth of the U.S. labor force was unemployed, and real GDP was 30 percent below its 1929 level.

This devastating episode caused many economists to question the validity of classical economic theory—the theory we examined in Chapters 3 through 7. Classical theory seemed incapable of explaining the Depression. According to that theory, national income depends on factor supplies and the available technology, neither of which changed substantially from 1929 to 1933. After the onset of the Depression, many economists believed that a new model was needed to explain such a large and sudden downturn and to suggest government policies that might reduce the economic hardship so many people faced.

In 1936 the British economist John Maynard Keynes revolutionized economics with his book *The General Theory of Employment, Interest, and Money*. Keynes proposed a new way to analyze the economy, which he presented as an alternative to classical theory. His vision of how the economy works quickly became a center of controversy. Yet as economists debated *The General Theory*, a new understanding of economic fluctuations gradually developed.

Keynes proposed that low aggregate demand is responsible for the low income and high unemployment that characterize economic downturns. He criticized classical theory for assuming that aggregate supply alone —capital, labor, and technology—determines national income. Economists today reconcile these views with the model of aggregate demand and aggregate supply introduced in Chapter 10. In the long run, prices are flexible, and aggregate supply determines income. But in the short run, prices are sticky, so changes in aggregate demand influence income.

Keynes's ideas about short-run fluctuations have been prominent since he proposed them in the 1930s, but they commanded renewed attention during the Great Recession of 2008–2009. As unemployment soared, policymakers debated how best to increase aggregate demand. Many of the issues that gripped economists during the Great Depression were once again at the center of the policy debate.

In this chapter and the next, we continue our study of economic fluctuations by looking more closely at aggregate demand. Our goal is to identify the variables that shift the aggregate demand curve, causing fluctuations in national income. We also examine more fully the tools policymakers can use to influence aggregate demand. In Chapter 10 we derived the aggregate demand curve from the quantity theory of money, and we showed that monetary policy can shift the aggregate demand curve. In this chapter we see that the government can influence aggregate demand with both monetary and fiscal policy.

The model of aggregate demand developed in this chapter, called the ***IS–LM* model**, is the leading interpretation of Keynes's theory. The goal of the model is to show what determines national income for a given price level. There are two ways to interpret this exercise. We can view the *IS–LM* model as showing what causes income to change in the short run when the price level is fixed because all prices are sticky. Or we can view the model as showing what causes the aggregate demand curve to shift. These two interpretations of the model are equivalent: as Figure 11-1 shows, in the short run when the price level is fixed, shifts in the aggregate demand curve lead to changes in the equilibrium level of national income.

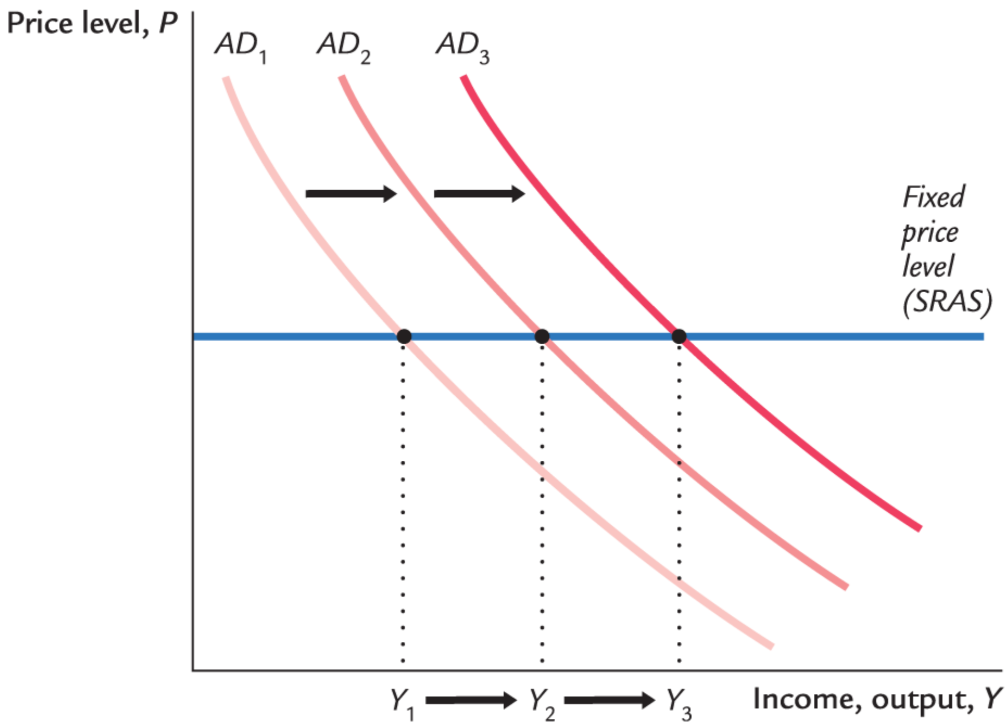


FIGURE 11-1 Shifts in Aggregate Demand For a given price level, national income fluctuates because of shifts in the aggregate demand curve. The *IS–LM* model takes the price level as given and shows what causes income to change. The model therefore shows what causes aggregate demand to shift.

The two parts of the *IS–LM* model are, not surprisingly, the ***IS* curve** and the ***LM* curve**. *IS* stands for "investment" and "saving," and the *IS* curve represents what's going on in the market for goods and services (which we first discussed in Chapter 3). *LM* stands for "liquidity" and "money," and the *LM* curve represents what's happening to the supply and demand for money (which we first discussed in Chapter 5). Because the interest rate influences both investment and money demand, it is the variable that links the two halves of the *IS–LM* model. The model shows how interactions between the goods and money markets determine the position and slope of the aggregate demand curve and, therefore, the level of national income in the short run.[1]

11-1 The Goods Market and the *IS* Curve

The *IS* curve plots the relationship between the interest rate and the level of income that arises in the market for goods and services. To develop this relationship, we start with a basic model called the **Keynesian cross**. This model is the simplest interpretation of Keynes's theory of how national income is determined and is a building block for the more complex and realistic *IS*–*LM* model.

The Keynesian Cross

In *The General Theory*, Keynes proposed that an economy's total income is, in the short run, determined largely by the spending plans of households, businesses, and government. The more people want to spend, the more goods and services firms can sell. The more firms can sell, the more output they will produce and the more workers they will hire. Keynes believed that the problem during recessions and depressions is inadequate spending. The Keynesian cross models this insight.

Planned Expenditure

We begin our derivation of the Keynesian cross by drawing a distinction between actual and planned expenditure. *Actual expenditure* is the amount households, firms, and the government spend on goods and services, and as we first saw in Chapter 2, it equals the economy's gross domestic product (GDP). *Planned expenditure* is the amount households, firms, and the government would like to spend on goods and services.

Why would actual expenditure ever differ from planned expenditure? The answer is that firms can have unplanned inventory investment when their sales do not meet their expectations. If firms sell less of their product than they planned, their stock of inventories automatically rises; conversely, if firms sell more than planned, their stock of inventories falls. Because these unplanned changes in inventory are counted as investment spending by firms, actual expenditure can be either above or below planned expenditure.

Now consider the determinants of planned expenditure. Assuming that the economy is closed, so that net exports are zero, we write planned expenditure *PE* as the sum of consumption *C*, planned investment *I*, and government purchases *G*:

$$PE = C + I + G.$$

To this equation, we add the consumption function:

$$C = C(Y - T).$$

This equation states that consumption depends on disposable income $(Y - T)$, which is total income *Y* minus taxes *T*. To keep things simple, for now we take planned investment as exogenously fixed:

$$I = \bar{I}.$$

Finally, as in Chapter 3, we assume that fiscal policy—the levels of government purchases and taxes—is fixed:

$$G = \bar{G}.$$
$$T = \bar{T}.$$

Combining these five equations, we obtain

$$PE = C(Y - \bar{T}) + \bar{I} + \bar{G}.$$

This equation shows that planned expenditure is a function of income *Y*, planned investment $\bar{I}$, and the fiscal policy variables $\bar{G}$ and $\bar{T}$.

Figure 11-2 graphs planned expenditure as a function of income. This line slopes upward because higher income leads to higher consumption and thus higher planned expenditure. The slope of this line is the marginal propensity to consume *MPC*: it shows how much planned expenditure increases when income rises by $1. This planned-expenditure function is the first piece of the Keynesian cross.

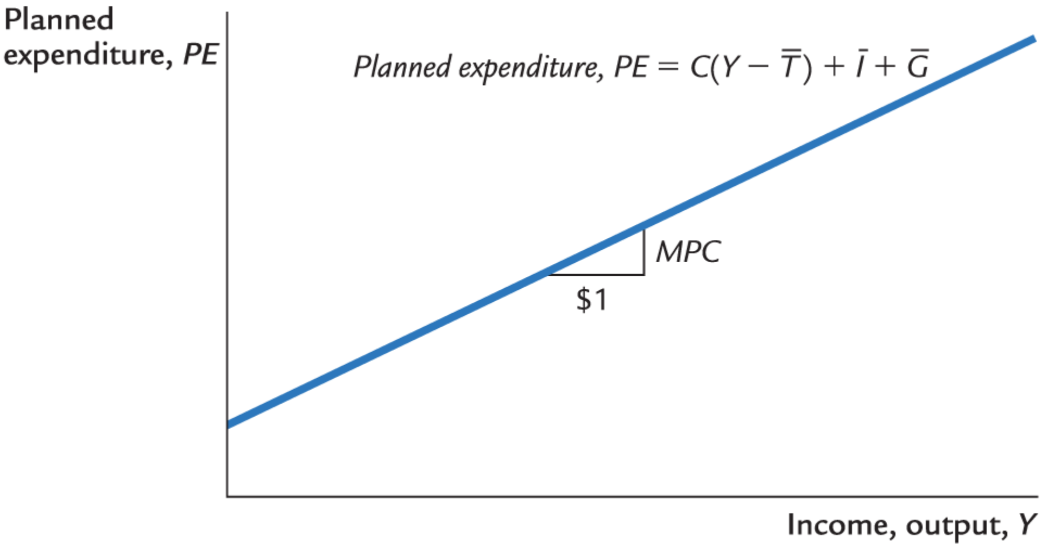


FIGURE 11-2 Planned Expenditure as a Function of Income Planned expenditure *PE* depends on income because higher income leads to higher consumption, which is part of planned expenditure. The slope of the planned-expenditure function is the marginal propensity to consume *MPC*.

The Economy in Equilibrium

The next piece of the Keynesian cross is the assumption that the economy is in equilibrium when actual expenditure equals planned expenditure. This assumption is based on the idea that when people's plans have been realized, they have no reason to change what they are doing. Recalling that *Y* as GDP equals not only total income but also total actual expenditure on goods and services, we can write this equilibrium condition as

$$\begin{aligned} \text{Actual Expenditure} &= \text{Planned Expenditure} \\ Y &= PE. \end{aligned}$$

Actual Expenditure=Planned ExpenditureY=PE.

The 45-degree line in Figure 11-3 plots the points where this condition holds. With the addition of the planned-expenditure function, this diagram becomes the Keynesian cross. The equilibrium of this economy is at point A, where the planned-expenditure function crosses the 45-degree line.

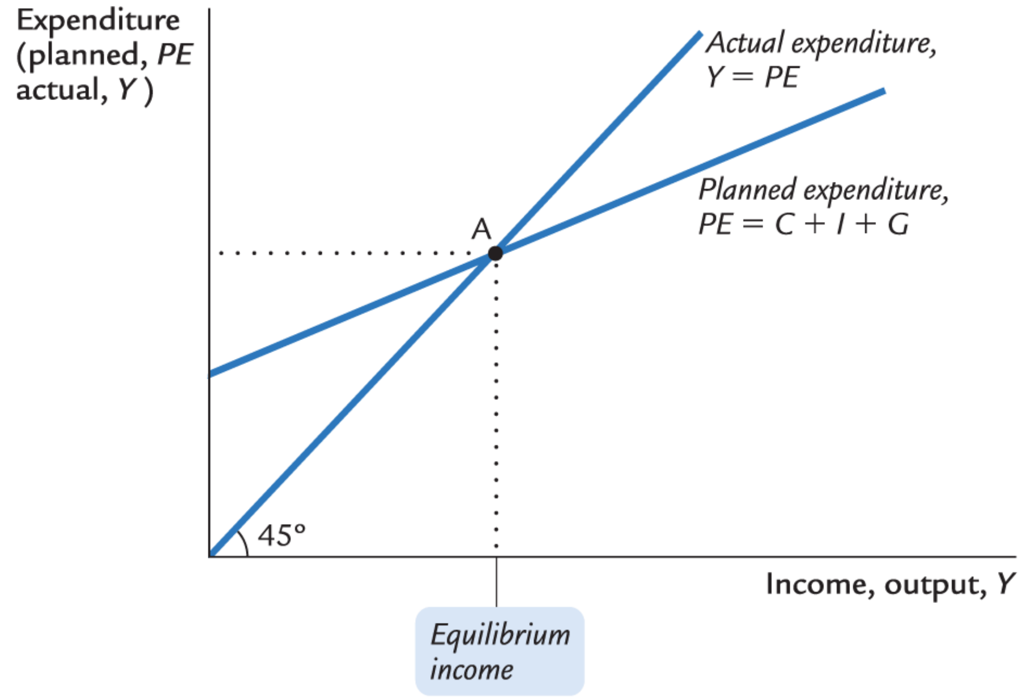
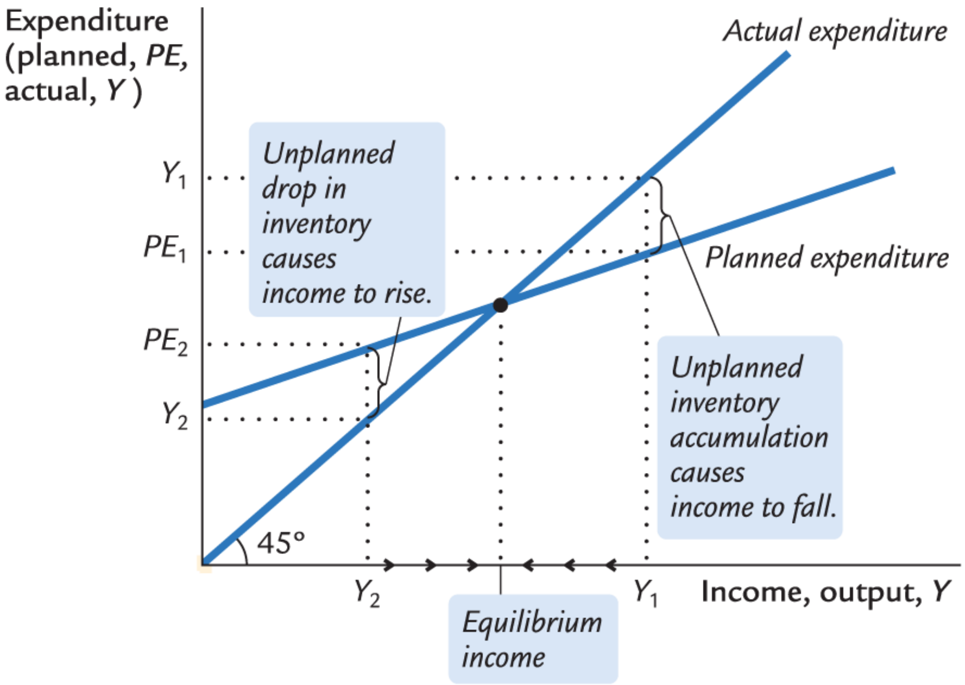


FIGURE 11-3 The Keynesian Cross The equilibrium in the Keynesian cross is the point at which income (actual expenditure) equals planned expenditure (point A).

How does the economy get to equilibrium? In this model, inventories play an important role in the adjustment process. Whenever an economy is not in equilibrium, firms experience unplanned changes in inventories, and this induces them to change production levels. Changes in production in turn influence total income and expenditure, moving the economy toward equilibrium.

For example, suppose the economy finds itself with GDP at a level greater than the equilibrium level, such as the level Y_1 in Figure 11-4. In this case, planned expenditure PE_1 is less than production Y_1, so firms are selling less than they are producing. Firms add the unsold goods to their stock of inventories. This unplanned rise in inventories prompts firms to lay off workers and cut production; these actions in turn reduce GDP. This process of unintended inventory accumulation and falling income continues until income *Y* falls to the equilibrium level.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

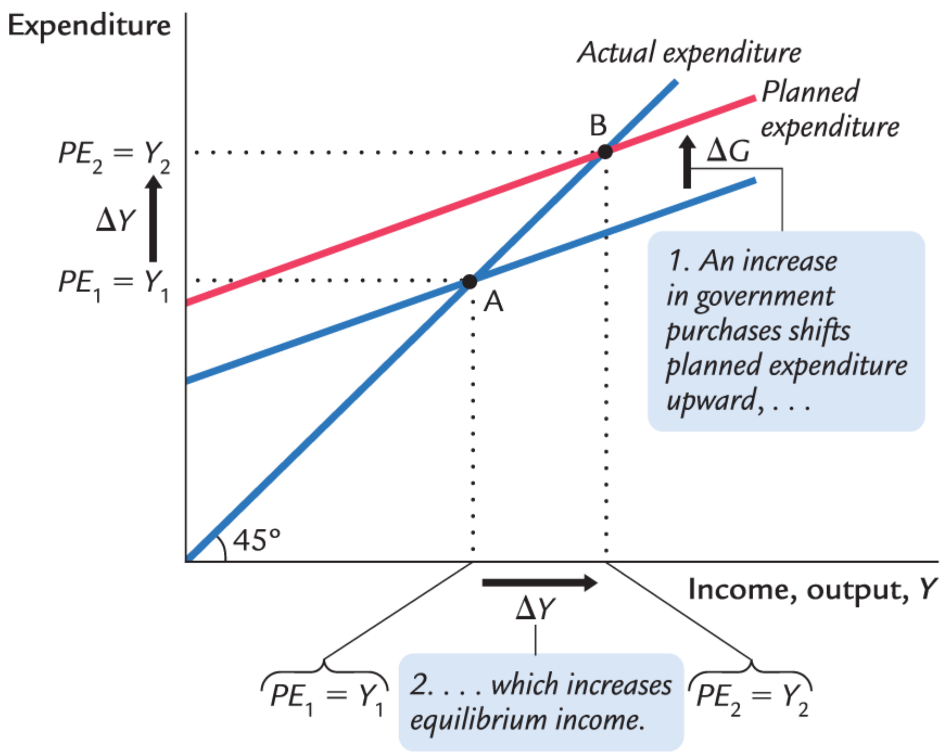
FIGURE 11-4 The Adjustment to Equilibrium in the Keynesian Cross If firms are producing at level Y_1, then planned expenditure PE_1 falls short of production, and firms accumulate inventories. This inventory accumulation induces firms to decrease production. Similarly, if firms are producing at level Y_2, then planned expenditure PE_2 exceeds production, and firms run down their inventories. This fall in inventories induces firms to increase production. In both cases, the firms' decisions drive the economy toward equilibrium.

Similarly, suppose GDP is at a level lower than the equilibrium level, such as the level Y_2 in Figure 11-4. In this case, planned expenditure PE_2 is greater than production Y_2. Firms satisfy customers by drawing down their inventories. But when firms see their stock of inventories dwindle, they hire more workers and increase production. GDP rises, and the economy approaches equilibrium.

In summary, the Keynesian cross shows how income *Y* is determined for given levels of planned investment *I* and fiscal policy *G* and *T*. We can use this model to show how income changes when one of these exogenous variables changes.

Fiscal Policy and the Multiplier: Government Purchases

Consider how changes in government purchases affect the economy. Because government purchases are one component of expenditure, higher government purchases result in higher planned expenditure for any given income. If government purchases rise by ΔG, then the planned-expenditure schedule shifts upward by ΔG, as in Figure 11-5. The equilibrium of the economy moves from point A to point B.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 11-5 An Increase in Government Purchases in the Keynesian Cross An increase in government purchases of ΔG raises planned expenditure by that amount for any given income. The equilibrium moves from point A to point B, and income rises from Y_1 to Y_2. Note that the increase in income ΔY exceeds the increase in government purchases ΔG. Thus, fiscal policy has a multiplied effect on income.

This graph shows that an increase in government purchases leads to an even greater increase in income. That is, ΔY is larger than ΔG. The ratio $\Delta Y/\Delta G$ is called the **government-purchases multiplier**; it tells us how much income rises in response to a $1 increase in government purchases. An implication of the Keynesian cross is that the government-purchases multiplier is larger than 1.

Why does fiscal policy have a multiplied effect on income? The reason is that, according to the consumption function $C = C(Y - T)$, higher income causes higher consumption. When an increase in government purchases raises income, it also raises consumption, which further raises income, which further raises consumption, and so on. Therefore, in this model, an increase in government purchases causes a greater increase in income.

How big is the multiplier? To answer this question, we trace through each step of the change in income. The process begins when expenditure rises by ΔG, implying that income rises by ΔG as well. This increase in income raises consumption by $MPC \times \Delta G$, where *MPC* is the marginal propensity to consume. This increase in consumption raises expenditure and income once again. This second increase in income of $MPC \times \Delta G$ again raises consumption, this time by $MPC \times (MPC \times \Delta G)$, which again raises expenditure and income, and so on. This feedback from consumption to income to consumption continues indefinitely. The total effect on income is

Initial Change in Government Purchases=ΔGFirst Change in Consumption=MPC ×ΔGSecond Change in C

$$\begin{array}{lcr}\text{Initial Change in Government Purchases} & = & \Delta G \\ \text{First Change in Consumption} & = & MPC \times \Delta G \\ \text{Second Change in Consumption} & = & MPC^2 \times \Delta G \\ \text{Third Change in Consumption} & = & MPC^3 \times \Delta G \\ . & & . \\ . & & . \\ . & & . \\ \hline \Delta Y = (1 + MPC + MPC^2 + MPC^3 + \ldots)\Delta G. & & \end{array}$$

The government-purchases multiplier is

ΔY/ΔG=1+MPC+MPC2+MPC3+....

$$\Delta Y/\Delta G = 1 + MPC + MPC^2 + MPC^3 + \ldots.$$

This expression for the multiplier is an example of an *infinite geometric series*. A result from algebra allows us to write the multiplier as[2]

ΔY/ΔG=1/(1−MPC).

$$\Delta Y/\Delta G = 1/(1 - MPC).$$

For example, if the marginal propensity to consume is 0.6, the multiplier is

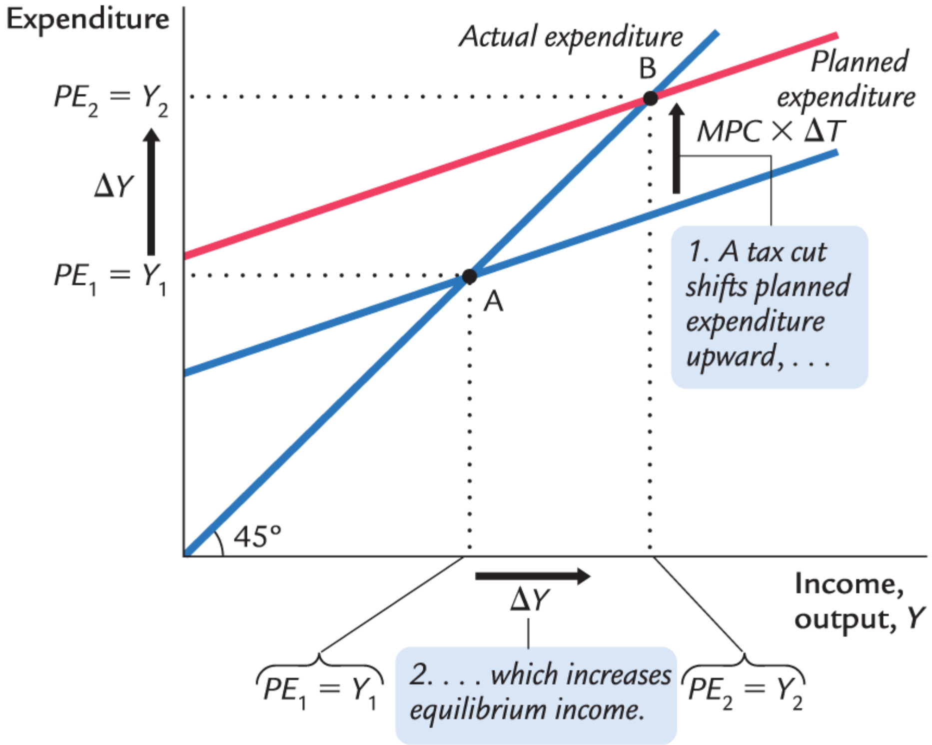
ΔY/ΔG=1+0.6+0.62+0.63+...=1/(1−0.6)=2.5.

$$\begin{aligned}\Delta Y/\Delta G &= 1 + 0.6 + 0.6^2 + 0.6^3 + \ldots \\ &= 1/(1 - 0.6) \\ &= 2.5.\end{aligned}$$

In this case, a $1.00 increase in government purchases raises equilibrium income by $2.50.[3]

Fiscal Policy and the Multiplier: Taxes

Now consider how changes in taxes affect equilibrium income. A decrease in taxes of ΔT immediately raises disposable income $Y - T$ by ΔT and, therefore, increases consumption by $MPC \times \Delta T$. For any given income *Y*, planned expenditure is now higher. As [Figure 11-6](#) shows, the planned-expenditure schedule shifts upward by $MPC \times \Delta T$. The equilibrium of the economy moves from point A to point B.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 11-6 A Decrease in Taxes in the Keynesian Cross A decrease in taxes of ΔT raises planned expenditure by $MPC \times \Delta T$ for any given income. The equilibrium moves from point A to point B, and income rises from Y_1 to Y_2. Again, fiscal policy has a multiplied effect on income.

Just as an increase in government purchases has a multiplied effect on income, so does a decrease in taxes. As before, the initial change in expenditure, now $MPC \times \Delta T$, is multiplied by $1/(1 - MPC)$. The overall effect on income of the change in taxes is

$$\Delta Y / \Delta T = -MPC / (1 - MPC).$$

This expression is the **tax multiplier**, the amount income changes in response to a $1 change in taxes. (The negative sign indicates that income moves in the opposite direction from taxes.) For example, if the marginal propensity to consume is 0.6, then the tax multiplier is

$$\Delta Y / \Delta T = -0.6 / (1 - 0.6) = -1.5.$$

In this example, a $1.00 cut in taxes raises equilibrium income by $1.50.[4]

CASE STUDY

Cutting Taxes to Stimulate the Economy: The Kennedy and Bush Tax Cuts

When John F. Kennedy became president of the United States in 1961, he recruited some of the brightest young economists of the day to work on his Council of Economic Advisers. These economists brought Keynesian ideas to discussions of economic policy at the highest level.

One of the council's first proposals was to expand national income by reducing taxes. This eventually led to a substantial cut in personal and corporate income taxes in 1964. The tax cut was intended to stimulate expenditure on consumption and investment and thus lead to higher levels of income and employment. When a reporter asked Kennedy why he advocated a tax cut, Kennedy replied, "To stimulate the economy. Don't you remember your Economics 101?"

As Kennedy's economic advisers predicted, the passage of the tax cut was followed by an economic boom. Growth in real GDP was 5.8 percent in 1964 and 6.5 percent in 1965. The unemployment rate fell from 5.6 percent in 1963 to 5.2 percent in 1964 and then to 4.5 percent in 1965.

Economists debate the source of this rapid growth in the early 1960s. A group called *supply-siders* argues that the economic boom resulted from the incentive effects of the cut in income tax rates. According to supply-siders, when workers can keep a higher fraction of their earnings, they supply substantially more labor and expand the aggregate supply of goods and services. Keynesians, however, emphasize the impact of tax cuts on aggregate demand. Most likely, both views contain some truth: *tax cuts stimulate aggregate supply by improving workers' incentives and expand aggregate demand by raising households' disposable income*.

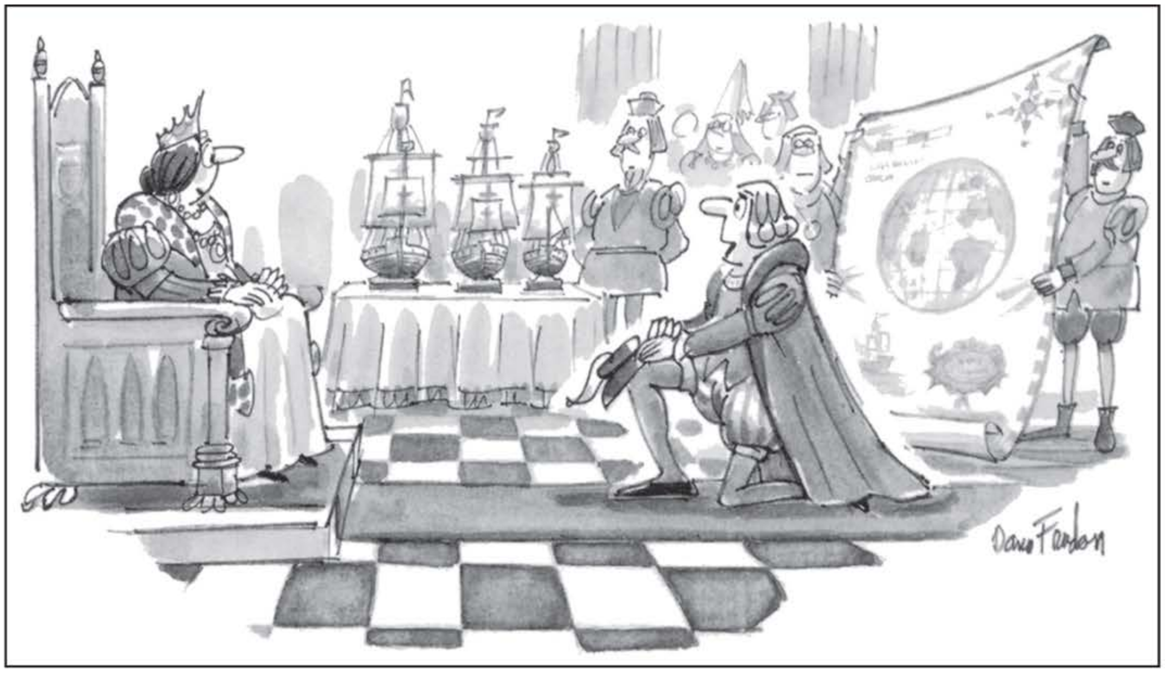
When George W. Bush was elected president in 2000, a big part of his platform was a cut in income taxes. Bush and his advisers used both supply-side and Keynesian rhetoric to make the case for their policy. (Disclosure: The author of this book was one of Bush's economic advisers from 2003 to 2005.) During the campaign, when the economy was doing fine, they argued that lower marginal tax rates would improve work incentives. But when the economy started to slow, and unemployment started to rise, the argument shifted to emphasize that the tax cut would stimulate spending and help the economy recover from the recession.

Congress passed major tax cuts in 2001 and 2003. After the second tax cut, the weak recovery from the 2001 recession turned into a more robust one. Growth in real GDP was 3.8 percent in 2004. The unemployment rate fell from its peak of 6.3 percent in June 2003 to 4.9 percent in December 2005.

When President Bush signed the 2003 tax bill, he explained the measure using the logic of aggregate demand: "When people have more money, they can spend it on goods and services. And in our society, when they demand an additional good or a service, somebody will produce the good or a service. And when somebody produces that good or a service, it means somebody is more likely to be able to find a job." The explanation could have come from an exam in Economics 101. ■

CASE STUDY

Increasing Government Purchases to Stimulate the Economy: The Obama Stimulus



Dana Fradon/The New Yorker/Conde Nast/The Cartoon Bank

“Your Majesty, my voyage will not only forge a new route to the spices of the East but also create over three thousand new jobs.”

When President Barack Obama took office in January 2009, the economy was suffering from a deep recession. (The recession’s causes are discussed in the next chapter and in more detail in Chapter 18.) Even before he was inaugurated, the president and his advisers proposed a sizable stimulus package to increase aggregate demand. As proposed, the package would cost the federal government about $800 billion, or about 5 percent of annual GDP. The package included some tax cuts and higher transfer payments, but much of it was made up of increases in government purchases of goods and services.

Professional economists debated the merits of the plan. Advocates of the Obama plan argued that increased spending was better than reduced taxes because, according to standard Keynesian theory, the government-purchases multiplier exceeds the tax multiplier. The reason for this difference is simple: when the government spends a dollar, that dollar gets spent, whereas when the government gives households a tax cut of a dollar, some of that dollar might be saved. According to an analysis by Obama administration economists, the government-purchases multiplier is 1.57, whereas the tax multiplier is only 0.99. Thus, they argued that increased government spending on roads, schools, and other infrastructure was the better route to increase aggregate demand and create jobs. The logic here is quintessentially Keynesian: as the economy sinks into recession, the government is acting as the demander of last resort.

The Obama stimulus proposal was controversial among economists for various reasons. One criticism was that the stimulus was not large enough, given the depth of the downturn. In March 2009, economist Paul Krugman wrote in the *New York Times*:

> The plan was too small and too cautious. . . . Employment has already fallen more in this recession than in the 1981–82 slump, considered the worst since the Great Depression. As a result, Mr. Obama’s promise that his plan will create or save 3.5 million jobs by the end of 2010 looks underwhelming, to say the least. It’s a credible promise—his economists used solidly mainstream estimates of the impacts of tax and spending policies. But 3.5 million jobs almost two years from now isn’t enough in the face of an economy that has already lost 4.4 million jobs, and is losing 600,000 more each month.

Still other economists argued that despite the predictions of conventional Keynesian models, spending-based fiscal stimulus is less effective than tax-based initiatives. A study of several dozen major countries since 1970 examined which kinds of fiscal stimulus have been most successful at promoting growth in economic activity. It found that fiscal stimulus is most successful when it primarily entails cuts in business and income taxes and least successful when it primarily entails increases in government spending.[5]

In addition, some economists thought that using infrastructure spending to promote employment might conflict with the goal of obtaining the infrastructure that was most needed. Here is how economist Gary Becker explained the concern on his blog:

> Putting new infrastructure spending in depressed areas like Detroit might have a big stimulating effect since infrastructure building projects in these areas can utilize some of the considerable unemployed resources there. However, many of these areas are also declining because they have been producing goods and services that are not in great demand, and will not be in demand in the future. Therefore, the overall value added by improving their roads and other infrastructure is likely to be a lot less than if the new infrastructure were located in growing areas that might have relatively little unemployment, but do have great demand for more roads, schools, and other types of long-term infrastructure.

In the end, Congress went ahead with President Obama's proposed stimulus with small modifications. The president signed the $787 billion bill on February 17, 2009. Did it work? The economy recovered from the recession, but more slowly than the Obama administration economists initially forecast. Whether the slow recovery reflects the failure of stimulus policy or a sicker economy than the economists first appreciated is a question of continuing debate. ■

CASE STUDY

Using Regional Data to Estimate Multipliers

As the preceding two case studies show, policymakers often change taxes and government spending to influence the economy. The short-run effects of these policy moves can be understood using Keynesian theory, such as the Keynesian cross and the *IS–LM* model. But do these policies work as well in practice as they do in theory?

Unfortunately, that question is hard to answer. When policymakers change fiscal policy, they usually do so for good reason. Because many other things are happening at the same time, there is no easy way to separate the effects of the fiscal policy from the effects of the other events. For example, President Obama proposed his 2009 stimulus plan because the economy was suffering in the aftermath of a financial crisis. We can observe what happened to the economy after the stimulus was passed, but disentangling the effects of the stimulus from the lingering effects of the financial crisis is a formidable task.

Increasingly, economists have tried to estimate multipliers for fiscal policy using regional data from states or provinces within a country. The use of regional data has two advantages. First, it increases the number of observations: the United States, for instance, has one national economy but 50 state economies. Second, and more important, it is possible to find variation in regional government spending that is plausibly unrelated to other events affecting the regional economy. By examining such random variation in government spending, a researcher can identify its economic effects without being led astray by other, confounding variables.

In one such study, Emi Nakamura and Jón Steinsson looked at the impact of defense spending on state economies. They began with the fact that states vary considerably in the size of their defense industries. For example, military contractors are more prevalent in California than in Illinois: when the U.S. federal government increases defense spending by 1 percent of U.S. GDP, defense spending in California rises on average by about 3 percent of California GDP, while defense spending in Illinois rises by only about 0.5 percent of Illinois GDP. By examining what happens to the California economy relative to the Illinois economy when the United States embarks on a military buildup, we can estimate the effects of government spending. Using data from all 50 states, Nakamura and Steinsson reported a government-purchases multiplier of 1.5. That is, when the government increases defense spending in a state by $1.00, it increases that state's GDP by $1.50.

In another study, Antonio Acconcia, Giancarlo Corsetti, and Saverio Simonelli used data from provinces within Italy to study the multiplier. Here the variation in government spending comes not from military buildups but from an Italian law cracking down on organized crime. According to the law, whenever the police uncover incriminating evidence that the Mafia has infiltrated a city council, the council is dismissed and replaced by external commissioners. These commissioners typically implement an immediate, unanticipated, and temporary cut in public investment projects. The study reported that this cut in government spending has a significant impact on the province's GDP. Once again, the multiplier is estimated to be about 1.5. Hence, these studies confirm the prediction of Keynesian theory that changes in government purchases can have a sizable effect on an economy's output of goods and services.

It is unclear, however, how to use these estimates from regional economies to draw inferences about national economies. One problem is that the regional government spending these researchers studied was not financed with regional taxes. Defense spending in California is largely paid for by federal taxes levied on the other 49 states, and the public investment projects in an Italian province are largely paid for at the national level. By contrast, when a nation increases its government spending, it must increase taxes, either in the present or the future, to pay for it. These higher taxes could depress economic activity, leading to a smaller multiplier. A second problem is that these regional changes in government spending do not influence monetary policy because central banks focus on national rather than regional conditions. By contrast, a national change in government spending could induce a change in monetary policy. In its attempt to stabilize the economy, the central bank may offset some of the effects of fiscal policy, making the multiplier smaller.

Although these two problems suggest that national multipliers are smaller than regional multipliers, a third problem works in the opposite direction. In a small regional economy, such as a state, many of the goods and services people buy are imported from neighboring states, whereas imports are a smaller share of a large national economy. When imports play a larger role, the marginal propensity to consume on domestic goods (those made within the state) is smaller. As the Keynesian cross describes, a smaller marginal propensity to consume on domestic goods leads to smaller second- and third-round effects and, thereby, a smaller multiplier. For this reason, national multipliers could be larger than regional multipliers.

The bottom line from studies of regional economies is that the demand from government purchases can exert a strong influence on economic activity. But the size of that effect as measured by the multiplier at the national level remains open to debate.[6] ■

The Interest Rate, Investment, and the *IS*

Curve

The Keynesian cross is only a stepping-stone on our path to the *IS–LM* model, which explains the economy's aggregate demand curve. The Keynesian cross is useful because it shows how the spending plans of households, firms, and the government determine the economy's income. Yet it makes the simplifying assumption that planned investment *I* is fixed. As we saw in Chapter 3, an important macroeconomic relationship is that planned investment depends on the interest rate *r*.

To add this relationship between the interest rate and investment to our model, we write planned investment as

$$I = I(r).$$

This investment function is graphed in panel (a) of Figure 11-7. Because the interest rate is the cost of borrowing to finance investment projects, an increase in the interest rate reduces planned investment. As a result, the investment function slopes downward.

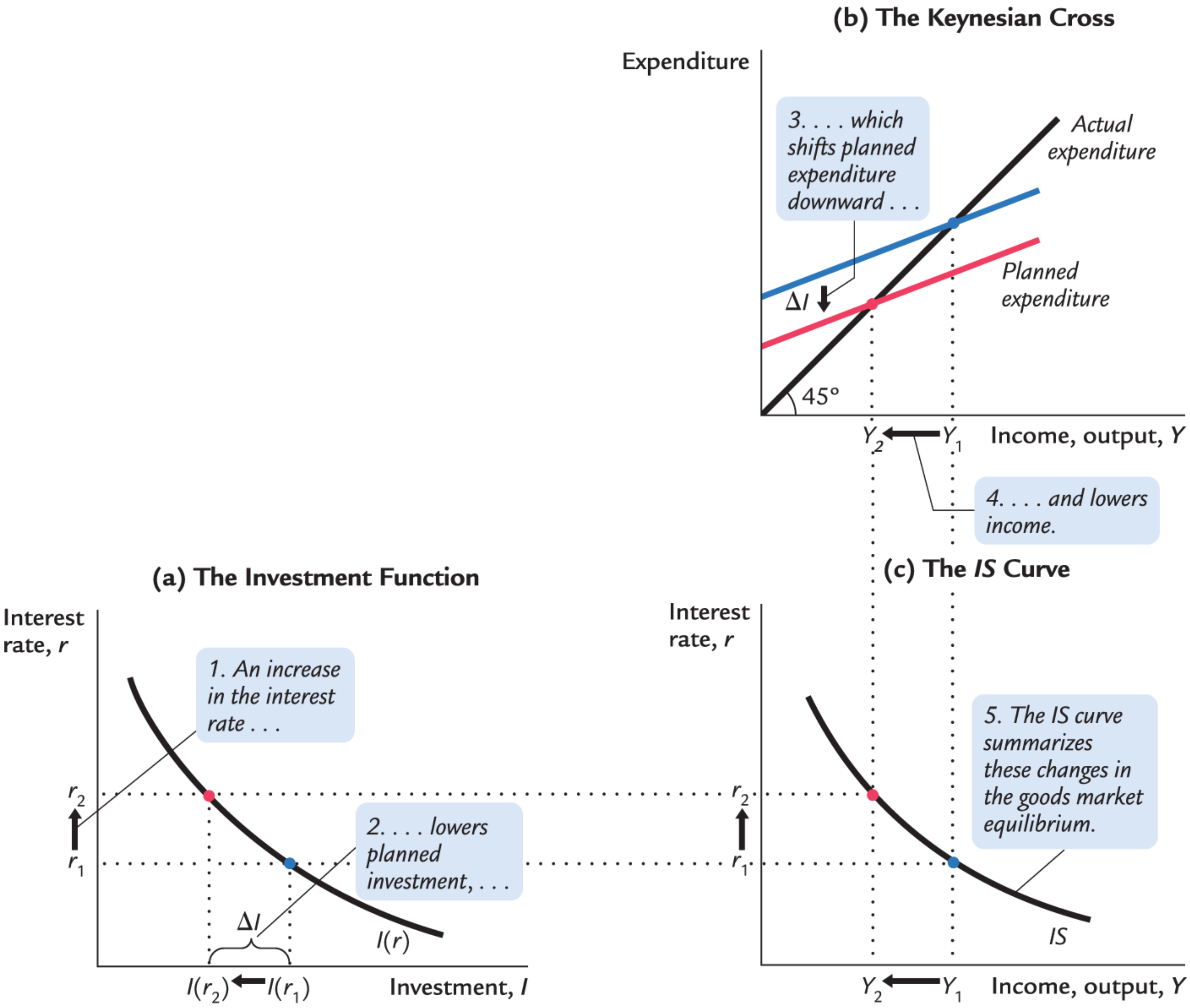


FIGURE 11-7 Deriving the *IS* Curve Panel (a) shows the investment function: an increase in the interest rate from r1 r_1 to r2 r_2 reduces planned investment from I(r1) $I(r_1)$ to I(r2). $I(r_2)$. Panel (b) shows the Keynesian cross: a decrease in planned investment from I(r1) $I(r_1)$ to I(r2) $I(r_2)$ shifts the planned-expenditure function downward, thereby reducing income from Y1 Y_1 to Y2. Y_2. Panel (c) shows the *IS* curve summarizing this relationship between the interest rate and income: the higher the interest rate, the lower the level of income.

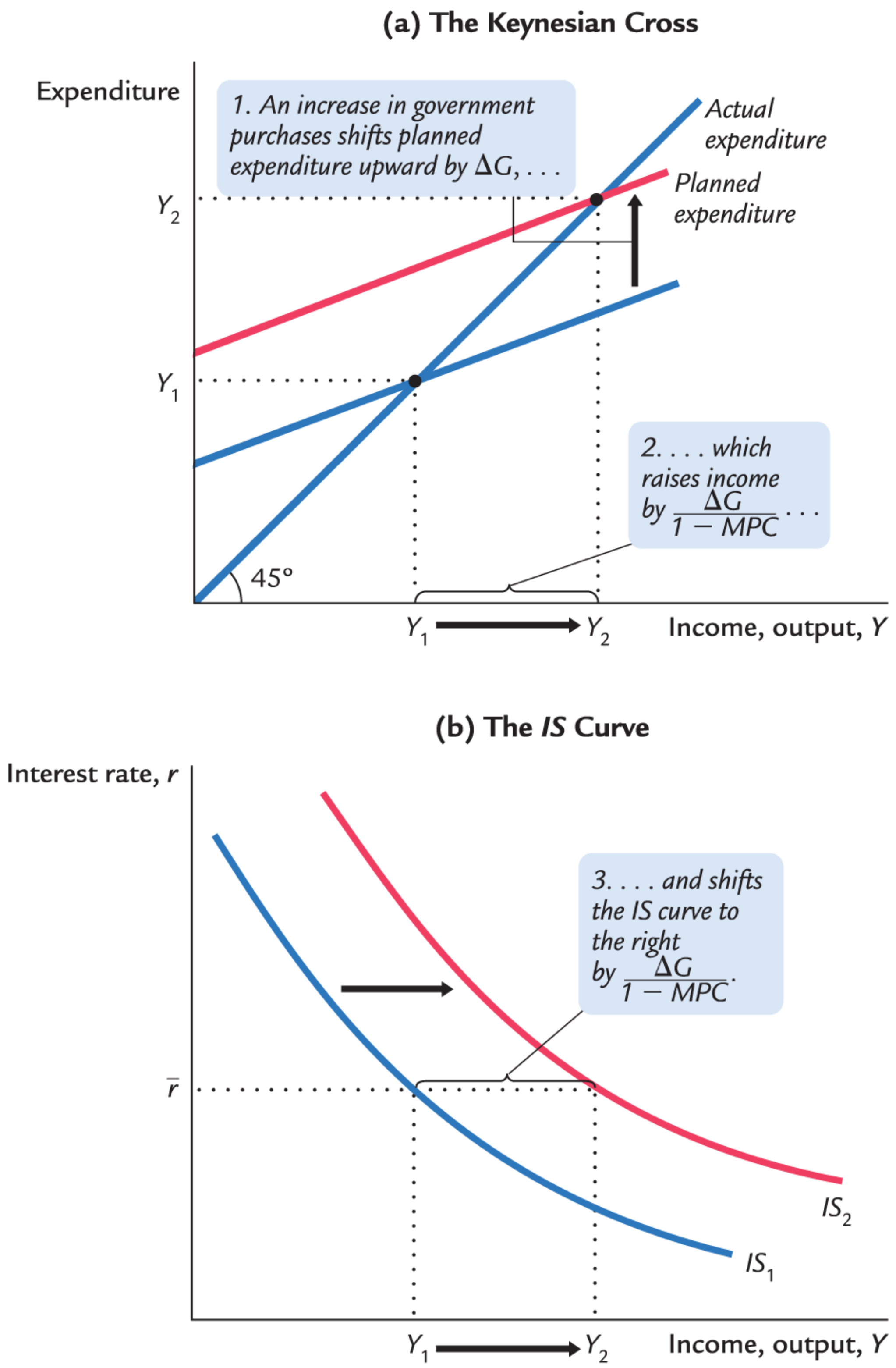
To determine how income changes when the interest rate changes, we can combine the investment function with the Keynesian-cross diagram. Because investment is inversely related to the interest rate, an increase in the interest rate from r1 r_1 to r2 r_2 reduces the quantity of investment from I(r1) $I(r_1)$ to I(r2). $I(r_2)$. The reduction in planned investment, in turn, shifts the planned-expenditure function downward, as in panel (b) of Figure 11-7. The shift in the planned-expenditure function causes income to fall from Y1 Y_1 to Y2. Y_2. Hence, an increase in the interest rate lowers income.

The *IS* curve, shown in panel (c) of Figure 11-7, summarizes this relationship between the interest rate and income. In essence, the *IS* curve combines the interaction between *r* and *I* expressed by the investment function and the interaction between *I* and *Y* demonstrated by the Keynesian cross. Each point on the *IS* curve represents equilibrium in the goods market, and the curve shows how equilibrium income depends on the interest rate. Because an increase in the interest rate causes planned investment to fall, which in turn causes income to fall, the *IS* curve slopes downward.

How Fiscal Policy Shifts the *IS* Curve

The *IS* curve shows us, for any given interest rate, the level of income that brings the goods market into equilibrium. As we learned from the Keynesian cross, equilibrium income also depends on government spending *G* and taxes *T*. The *IS* curve is drawn for a given fiscal policy; that is, when we construct the *IS* curve, we hold *G* and *T* fixed. When fiscal policy changes, the *IS* curve shifts.

Figure 11-8 uses the Keynesian cross to show how an increase in government purchases ΔG shifts the *IS* curve. This figure is drawn for a given interest rate $\bar{r}$ and thus for a given level of planned investment. The Keynesian cross in panel (a) shows that this change in fiscal policy raises planned expenditure and thereby increases equilibrium income from Y_1 to Y_2. Therefore, in panel (b), the increase in government purchases shifts the *IS* curve outward.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 11-8 An Increase in Government Purchases Shifts the *IS* Curve Outward Panel (a) shows that an increase in government purchases raises planned expenditure. For any given interest rate, the upward shift in planned expenditure of ΔGΔG leads to an increase in income *Y* of ΔG/(1−MPC).$\Delta G/(1-MPC)$. Therefore, in panel (b), the *IS* curve shifts to the right by this amount.

We can use the Keynesian cross to see how other changes in fiscal policy shift the *IS* curve. Because a decrease in taxes also expands expenditure and income, it shifts the *IS* curve outward as well. A decrease in government purchases or an increase in taxes reduces income; therefore, such a change in fiscal policy shifts the *IS* curve inward.

In summary, the IS *curve shows the combinations of the interest rate and income that are consistent with equilibrium in the market for goods and services. The* IS *curve is drawn for a given fiscal policy. Changes in fiscal policy that raise the demand for goods and services shift the* IS *curve to the right. Changes in fiscal policy that reduce the demand for goods and services shift the* IS *curve to the left.*

11-2 The Money Market and the *LM* Curve

The *LM* curve plots the relationship between the interest rate and the level of income that arises in the market for money balances. To understand this relationship, we begin by looking at a theory of the interest rate called the **theory of liquidity preference**.

The Theory of Liquidity Preference

In his classic work *The General Theory*, Keynes offered his view of how the interest rate is determined in the short run. His explanation is called the theory of liquidity preference because it posits that the interest rate adjusts to balance the supply and demand for the economy’s most liquid asset—money. Just as the Keynesian cross is a building block for the *IS* curve, the theory of liquidity preference is a building block for the *LM* curve.

To develop this theory, we begin with the supply of real money balances. If *M* stands for the supply of money and *P* stands for the price level, then M/P is the supply of real money balances. The theory of liquidity preference assumes there is a fixed supply of real money balances. That is,

$$(M/P)^s = \overline{M}/\overline{P}.$$

The money supply *M* is an exogenous policy variable chosen by a central bank, such as the Federal Reserve. The price level *P* is also an exogenous variable in this model. (We take the price level as given because the *IS*–*LM* model explains the short run when the price level is fixed.) These assumptions imply that the supply of real money balances is fixed and, in particular, does not depend on the interest rate. Thus, when we plot the supply of real money balances against the interest rate in Figure 11-9, we obtain a vertical supply curve.

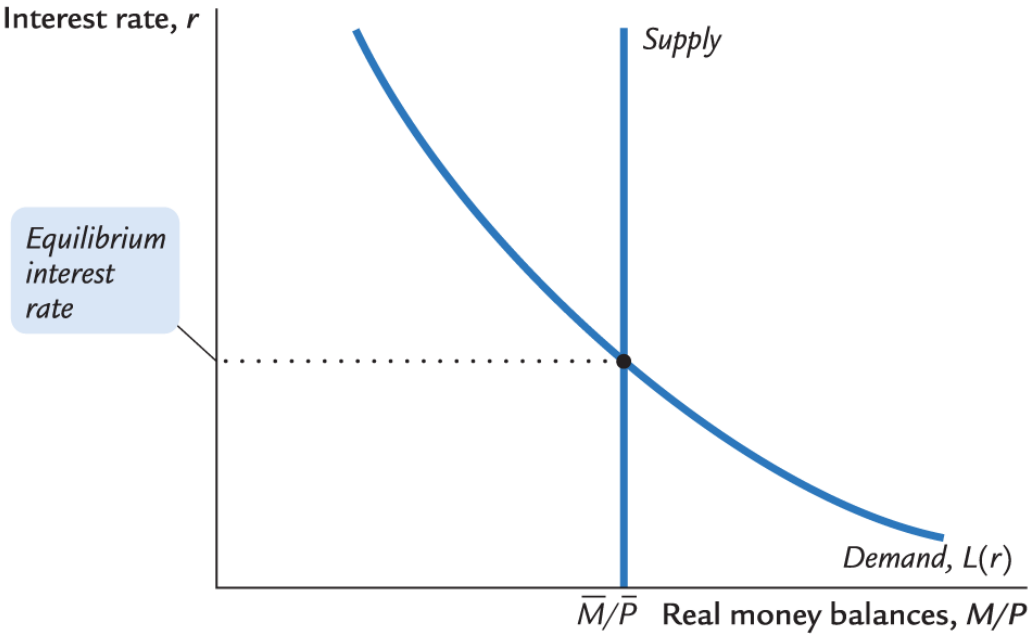


FIGURE 11-9 The Theory of Liquidity Preference The supply and demand for real money balances determine the interest rate. The supply curve for real money balances is vertical because the quantity of real money balances supplied does not depend on the interest rate. The demand curve slopes downward because a higher interest rate raises the cost of holding money and thus lowers the quantity demanded. At the equilibrium interest rate, the quantity of real money balances demanded equals the quantity supplied.

Next, consider the demand for real money balances. The theory of liquidity preference posits that the interest rate is one determinant of how much money people choose to hold. The underlying reason is that the interest rate is the opportunity cost of holding money: it is what you forgo by holding some of your assets as money, which does not bear interest, instead of as interest-bearing bank deposits or bonds. When the interest rate rises, people want to hold less of their wealth in the form of money. We can write the demand for real money balances as

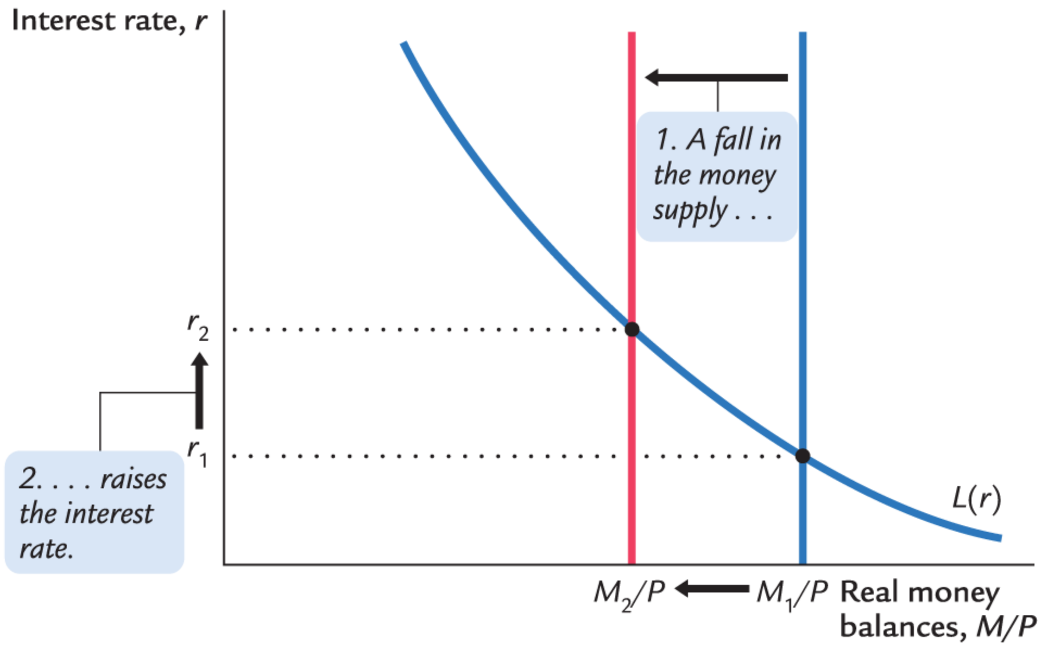
$$(M/P)^d = L(r),$$

where the function $L(\ )$ shows that the quantity of money demanded depends on the interest rate. The demand curve in Figure 11-9 slopes downward because higher interest rates reduce the quantity of real money balances demanded.[7]

According to the theory of liquidity preference, the supply and demand for real money balances determine what interest rate prevails in the economy. That is, the interest rate adjusts to equilibrate the money market. As the figure shows, at the equilibrium interest rate, the quantity of real money balances demanded equals the quantity supplied.

How does the interest rate get to this equilibrium of money supply and money demand? The adjustment occurs because whenever the money market is not in equilibrium, people try to adjust their portfolios of assets and, in the process, alter the interest rate. For instance, if the interest rate is above the equilibrium level, the quantity of real money balances supplied exceeds the quantity demanded. Individuals holding the excess supply of money try to convert some of their non-interest-bearing money into interest-bearing bank deposits or bonds. Banks and bond issuers, which prefer to pay lower interest rates, respond to this excess supply of money by lowering the interest rates they offer. Conversely, if the interest rate is below the equilibrium level, so that the quantity of money demanded exceeds the quantity supplied, individuals try to obtain money by selling bonds or making bank withdrawals. To attract now-scarcer funds, banks and bond issuers respond by increasing the interest rates they offer. Eventually, the interest rate reaches the equilibrium level, at which people are content with their portfolios of monetary and nonmonetary assets.

Now that we know how the interest rate is determined, we can use the theory of liquidity preference to show how the interest rate responds to changes in the supply of money. Suppose, for instance, the Fed decreases the money supply. A fall in *M* reduces M/P because *P* is fixed. The supply of real money balances shifts to the left, as in Figure 11-10. The equilibrium interest rate rises from r_1 to r_2, and the higher interest rate makes people satisfied to hold the smaller quantity of real money balances. The opposite occurs if the Fed increases the money supply. Thus, according to the theory of liquidity preference, a decrease in the money supply raises the interest rate, and an increase in the money supply lowers the interest rate.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 11-10 A Reduction in the Money Supply in the Theory of Liquidity Preference If the price level is fixed, a reduction in the money supply from M_1 to M_2 reduces the supply of real money balances. The equilibrium interest rate therefore rises from r_1 to r_2.

CASE STUDY

Does a Monetary Tightening Raise or Lower Interest Rates?

How does a tightening of monetary policy influence nominal interest rates? According to the theories we have been developing, the answer depends on the time horizon. Our analysis of the Fisher effect in Chapter 5 suggests that, in the long run when prices are flexible, a reduction in money growth would lower inflation, resulting in lower nominal interest rates. Yet the theory of liquidity preference predicts that, in the short run when prices are sticky, anti-inflationary monetary policy would lead to falling real money balances and higher interest rates.

Both conclusions are consistent with experience. A good example occurred during the early 1980s, when the U.S. economy saw a large and quick reduction in inflation.

Here's the background: by the late 1970s, inflation in the U.S. economy had reached the double-digit range and was a major national problem. In 1979 consumer prices were rising at a rate of 11.3 percent per year. In October of that year, only two months after becoming Fed chair, Paul Volcker decided that it was time to change course. He announced that monetary policy would aim to reduce inflation. This announcement began a period of tight money that, by 1983, brought the inflation rate down to 3.2 percent.

Let's look at what happened to nominal interest rates. In the period immediately after the October 1979 announcement of tighter monetary policy, we see a fall in real money balances and a rise in the interest rate—just as the theory of liquidity preference predicts. Nominal interest rates on three-month Treasury bills rose from 10.3 percent just before the October 1979 announcement to 11.4 percent in 1980 and 14.0 percent in 1981. Yet these high interest rates were only temporary. As Volcker's change in monetary policy lowered inflation and expectations of inflation, nominal interest rates gradually fell, reaching 6.0 percent in 1986.

This episode illustrates a general lesson: to understand the link between monetary policy and nominal interest rates, we need to keep in mind both the theory of liquidity preference and the Fisher effect. A monetary tightening leads to higher nominal interest rates in the short run and lower nominal interest rates in the long run. ■

Income, Money Demand, and the *LM* Curve

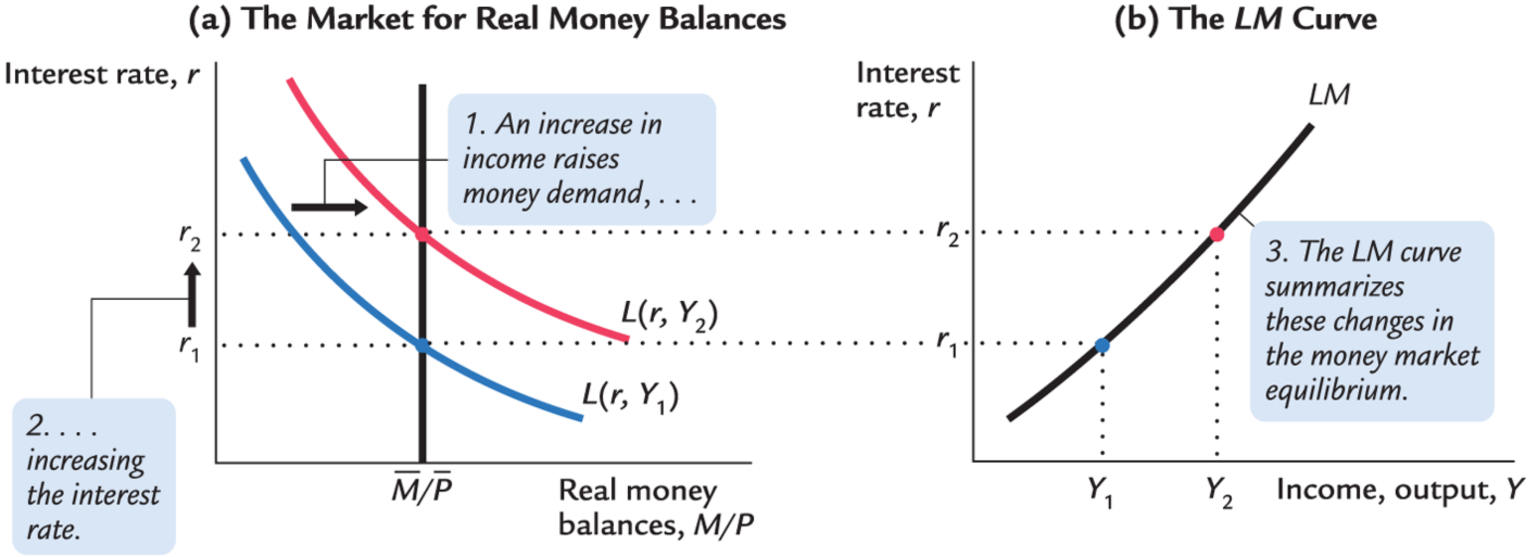
Having developed the theory of liquidity preference as an explanation for how the interest rate is determined, we can now use the theory to derive the *LM* curve. We begin by considering the following question: How does a change in the economy's income *Y* affect the market for real money balances? The answer (which should be familiar from Chapter 5) is that income affects the demand for money. When income is high, expenditure is high, and people engage in more transactions that require the use of money. Thus, greater income implies greater money demand. We can express these ideas by writing the money demand function as

$$(M/P)^d = L(r, Y).$$

The quantity of real money balances demanded is negatively related to the interest rate and positively related to income.

Using the theory of liquidity preference, we can figure out what happens to the equilibrium interest rate when income changes. For example, consider what happens in Figure 11-11 when income increases from Y_1

Y_1 to Y_2. As panel (a) illustrates, this increase in income shifts the money demand curve to the right. With the supply of real money balances unchanged, the interest rate must rise from r_1 to r_2 to equilibrate the money market. Therefore, according to the theory of liquidity preference, higher income leads to a higher interest rate.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

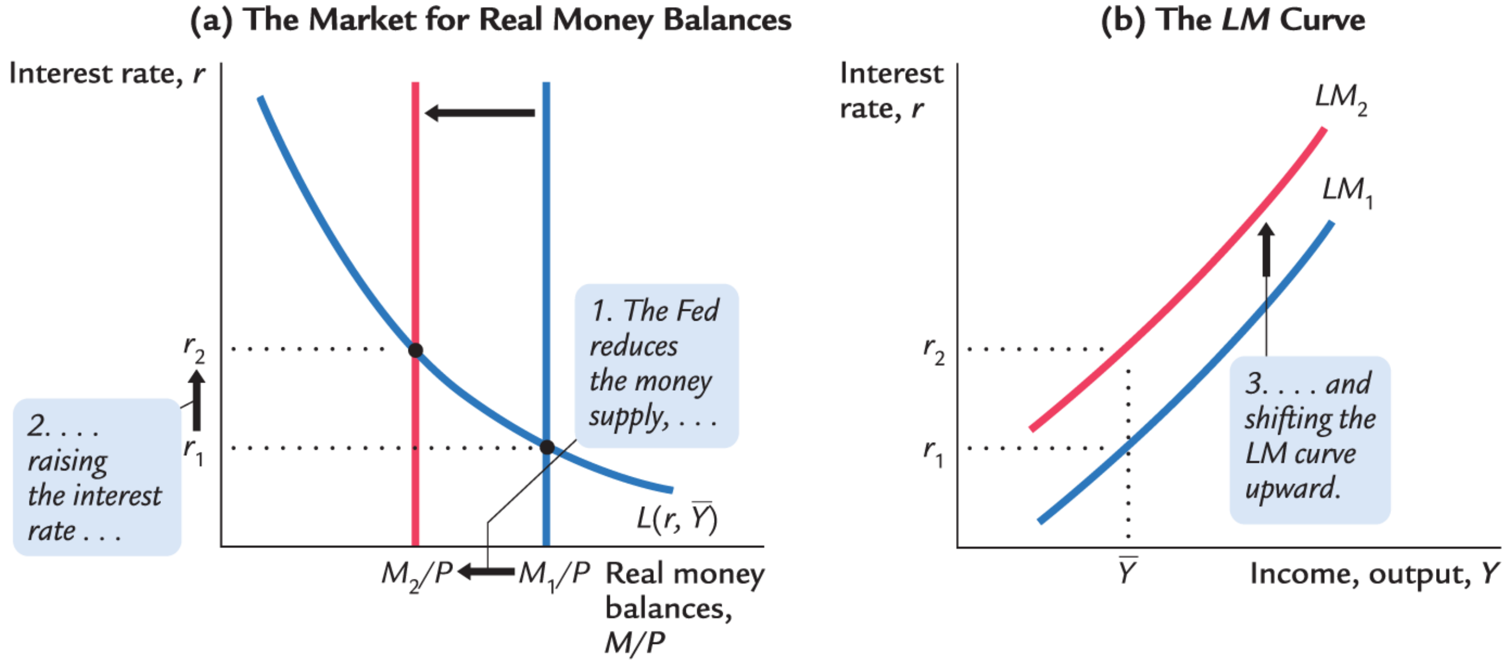
FIGURE 11-11 Deriving the *LM* Curve Panel (a) shows the market for real money balances: an increase in income from Y_1 to Y_2 raises the demand for money and thus raises the interest rate from r_1 to r_2. Panel (b) shows the *LM* curve summarizing this relationship between the interest rate and income: the higher the level of income, the higher the interest rate.

The *LM* curve shown in panel (b) of Figure 11-11 summarizes this relationship between income and the interest rate. Each point on the *LM* curve represents equilibrium in the money market, and the curve shows how the equilibrium interest rate depends on income. The higher the level of income, the higher the demand for real money balances, and the higher the equilibrium interest rate. For this reason, the *LM* curve slopes upward.

How Monetary Policy Shifts the *LM* Curve

The *LM* curve tells us the interest rate that equilibrates the money market at any level of income. Yet, as we saw earlier, the equilibrium interest rate also depends on the supply of real money balances M/P. This means that the *LM* curve is drawn for a *given* supply of real money balances. If real money balances change—for example, if the Fed alters the money supply—the *LM* curve shifts.

We can use the theory of liquidity preference to understand how monetary policy shifts the *LM* curve. Suppose the Fed decreases the money supply from M_1 to M_2, causing the supply of real money balances to fall from M_1/P to M_2/P. Figure 11-12 shows what happens. Holding constant income and thus the demand curve for real money balances, we see that a reduction in the supply of real money balances raises the interest rate that equilibrates the money market. Hence, a decrease in the money supply shifts the *LM* curve upward.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 11-12 A Reduction in the Money Supply Shifts the *LM* Curve Upward Panel (a) shows that for any given income Y‾, $\overline{Y}$, a reduction in the money supply raises the interest rate that equilibrates the money market. Therefore, the *LM* curve in panel (b) shifts upward.

In summary, the LM *curve shows the combinations of the interest rate and income that are consistent with equilibrium in the market for real money balances. The* LM *curve is drawn for a given supply of real money balances. Decreases in the supply of real money balances shift the* LM *curve upward. Increases in the supply of real money balances shift the* LM *curve downward.*

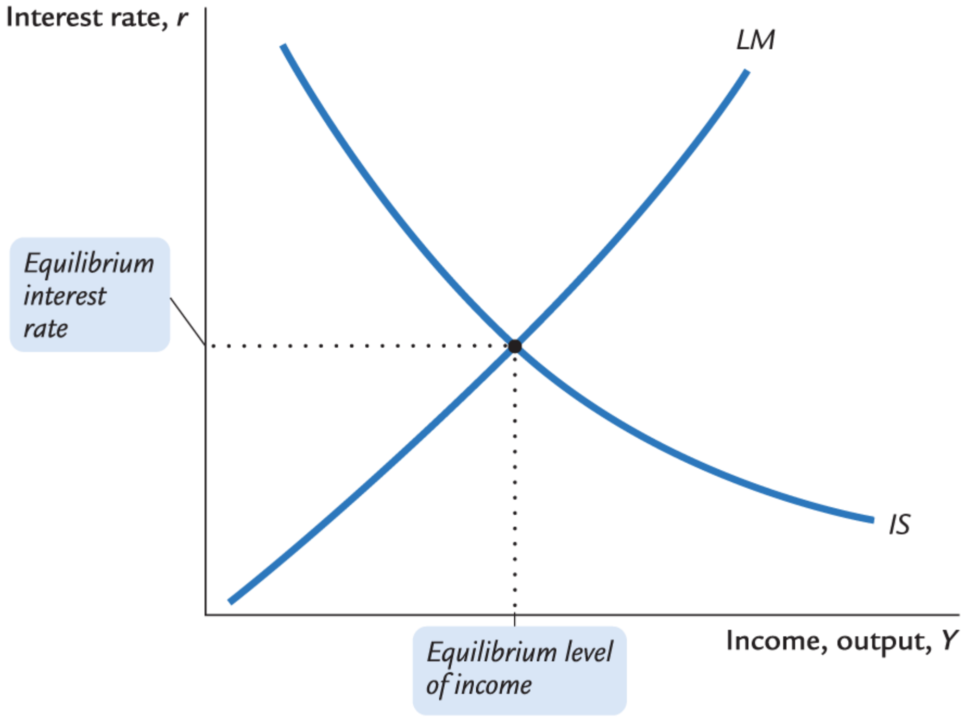
11-3 Conclusion: The Short-Run Equilibrium

We now have all the pieces of the *IS*–*LM* model. The two equations of this model are

Y=C(Y−T)+I(r)+GIS,M/P=L(r, Y)LM.

$$\begin{aligned} Y &= C(Y-T)+I(r)+G \quad && IS, \\ M/P &= L(r,\ Y) && LM. \end{aligned}$$

The model takes fiscal policy *G* and *T*, monetary policy *M*, and the price level *P* as exogenous. Given these exogenous variables, the *IS* curve provides the combinations of *r* and *Y* that satisfy the equation representing the goods market, and the *LM* curve provides the combinations of *r* and *Y* that satisfy the equation representing the money market. These two curves are shown together in Figure 11-13.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 11-13 Equilibrium in the *IS*–*LM* Model The intersection of the *IS* and *LM* curves represents simultaneous equilibrium in the market for goods and services and in the market for real money balances for given values of government spending, taxes, the money supply, and the price level.

The equilibrium of the economy is the point at which the *IS* curve and the *LM* curve cross. This point gives the interest rate *r* and income *Y* that satisfy conditions for equilibrium in both the goods market and the money market. In other words, at this intersection, actual expenditure equals planned expenditure, and the demand for real money balances equals the supply.

As we conclude this chapter, let’s recall that our goal in developing the *IS–LM* model is to analyze short-run fluctuations in economic activity. Figure 11-14 shows how the different pieces of our theory fit together. In this chapter we developed the Keynesian cross and the theory of liquidity preference as building blocks for the *IS–LM* model. As we discuss in the next chapter, the *IS–LM* model helps explain the position and slope of the aggregate demand curve. The aggregate demand curve, in turn, is a piece of the model of aggregate supply and aggregate demand, which economists use to explain the short-run effects of policy changes and other events on national income.

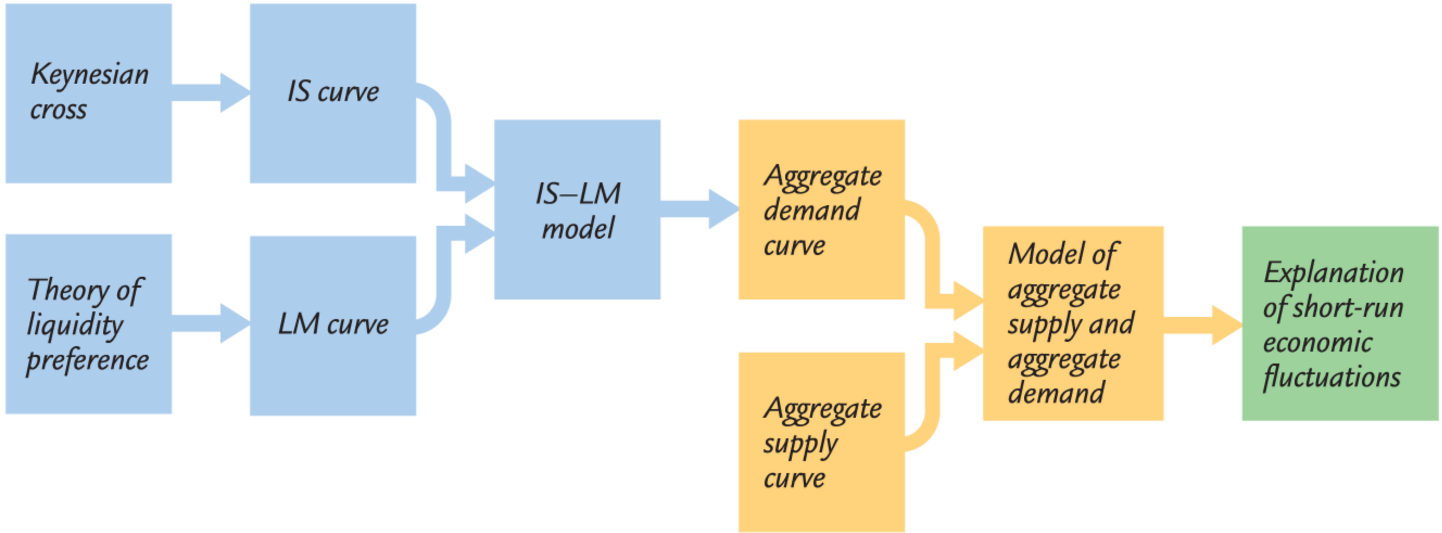


FIGURE 11-14 The Theory of Short-Run Fluctuations This schematic diagram shows how the different pieces of the theory of short-run fluctuations fit together. The Keynesian cross explains the *IS* curve, and the theory of liquidity preference explains the *LM* curve. The *IS* and *LM* curves together yield the *IS–LM* model, which explains the aggregate demand curve. The aggregate demand curve is part of the model of aggregate supply and aggregate demand. Economists use this model to explain short-run fluctuations in economic activity.

CHAPTER 12

Aggregate Demand II: Applying the *IS–LM* Model

Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

> Science is a parasite: the greater the patient population the better the advance in physiology and pathology; and out of pathology arises therapy. The year 1932 was the trough of the Great Depression, and from its rotten soil was belatedly begot a new subject that today we call macroeconomics.
>
> —Paul Samuelson

In Chapter 11 we assembled the pieces of the *IS–LM* model as a step toward understanding short-run economic fluctuations. We saw that the *IS* curve represents the equilibrium in the market for goods and services, that the *LM* curve represents the equilibrium in the market for real money balances, and that the *IS* and *LM* curves together determine the interest rate and national income in the short run when the price level is fixed. Now we turn our attention to applying the *IS–LM* model to analyze three issues.

First, we examine the potential causes of fluctuations in national income. We use the *IS–LM* model to see how changes in the exogenous variables (government purchases, taxes, and the money supply) influence the endogenous variables (the interest rate and national income) for a given price level. We also examine how various shocks to the goods market (the *IS* curve) and the money market (the *LM* curve) affect the interest rate and national income in the short run.

Second, we discuss how the *IS–LM* model fits into the model of aggregate supply and aggregate demand we introduced in Chapter 10. In particular, we examine how the *IS–LM* model provides a theory to explain the slope and position of the aggregate demand curve. Here we relax the assumption that the price level is fixed and show that the *IS–LM* model implies a negative relationship between the price level and national income. The model also reveals what events shift the aggregate demand curve and in what direction.

Third, we examine the Great Depression of the 1930s. As this chapter's opening quotation indicates, this episode gave birth to short-run macroeconomic theory, for it led Keynes and his many followers to argue that aggregate demand is the key to understanding fluctuations in national income. With the benefit of hindsight, we can use the *IS–LM* model to discuss the various explanations of this traumatic economic downturn.

The *IS–LM* model has played a central role in the history of economic thought, and it offers a powerful lens through which to view economic history, but it has much modern significance as well. Throughout this chapter we will see that the model can also be used to shed light on more recent fluctuations in the economy; two case studies in the chapter use it to examine the recessions that began in 2001 and 2008. Moreover, as we will see in Chapter 15, the logic of the *IS–LM* model provides the foundation for understanding newer and more sophisticated theories of the business cycle.

12-1 Explaining Fluctuations with the *IS–LM* Model

The intersection of the *IS* curve and the *LM* curve determines national income. When one of these curves shifts, the short-run equilibrium of the economy changes, and national income fluctuates. In this section we examine how changes in policy and shocks to the economy can cause these curves to shift.

How Fiscal Policy Shifts the *IS* Curve and Changes the Short-Run Equilibrium

Let’s first examine how changes in fiscal policy (government purchases and taxes) affect the economy’s short-run equilibrium. Recall that changes in fiscal policy influence planned expenditure and thereby shift the *IS* curve. The *IS–LM* model shows how these shifts in the *IS* curve affect income and the interest rate.

Changes in Government Purchases

Consider an increase in government purchases of ΔG. The government-purchases multiplier in the Keynesian cross tells us that this change in fiscal policy raises income at any given interest rate by $\Delta G/(1-MPC)$. Therefore, as Figure 12-1 shows, the *IS* curve shifts to the right by this amount. The equilibrium of the economy moves from point A to point B. The increase in government purchases raises both income and the interest rate.

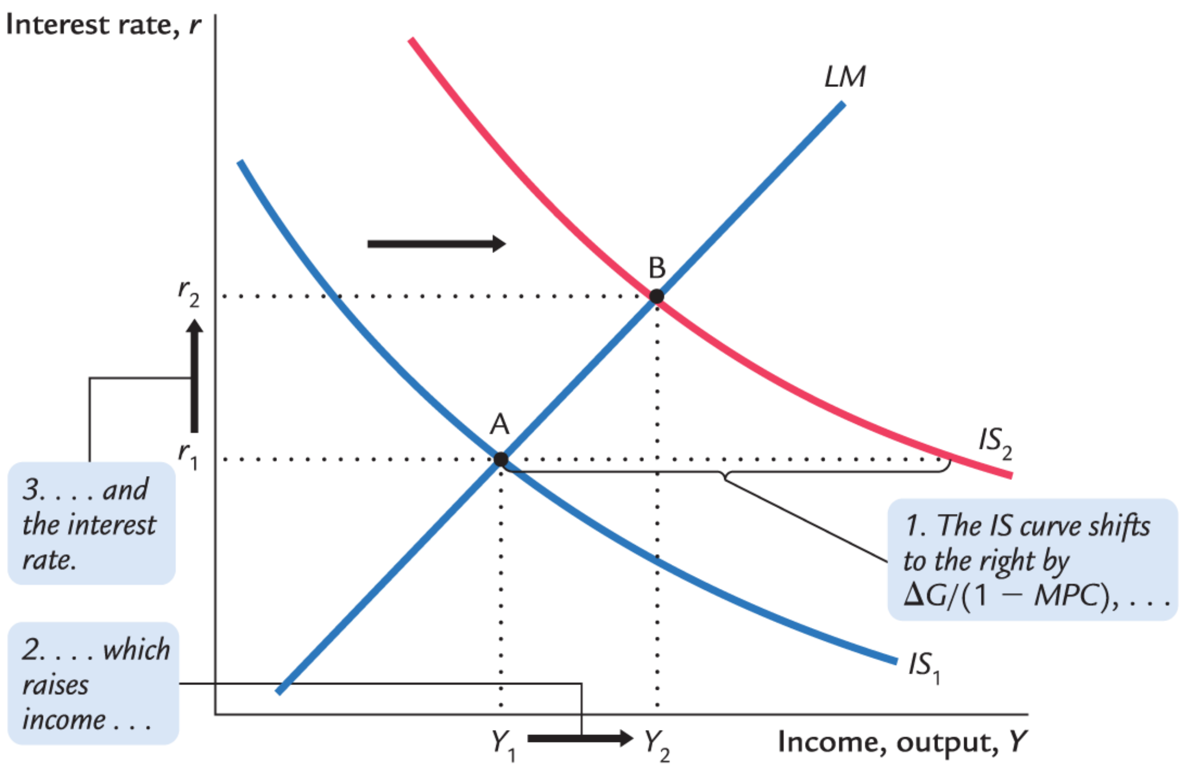


FIGURE 12-1 An Increase in Government Purchases in the *IS–LM* Model An increase in government purchases shifts the *IS* curve to the right. The equilibrium moves from point A to point B. Income rises from Y1Y_1 to Y2,Y_2, and the interest rate rises from r1r_1 to r2.r_2.

To understand fully what's happening in Figure 12-1, it helps to keep in mind the building blocks for the *IS–LM* model from the preceding chapter—the Keynesian cross and the theory of liquidity preference. Here is the story. When the government increases its purchases of goods and services, the economy's planned expenditure rises. The increase in planned expenditure stimulates the production of goods and services, which causes total income *Y* to rise. These effects should be familiar from the Keynesian cross.

Now consider the money market, as described by the theory of liquidity preference. Because the economy's demand for money depends on income, the rise in total income increases the quantity of money demanded at every interest rate. The supply of money, however, has not changed, so higher money demand causes the equilibrium interest rate *r* to rise.

The higher interest rate arising in the money market, in turn, has ramifications in the goods market. When the interest rate rises, firms cut back on their investment plans. This fall in investment partially offsets the expansionary effect of the increase in government purchases. Thus, the increase in income in response to a fiscal expansion is smaller in the *IS–LM* model than it is in the Keynesian cross (where investment is assumed to be fixed). You can see this in Figure 12-1. The horizontal shift in the *IS* curve equals the rise in equilibrium income in the Keynesian cross. This amount is larger than the increase in equilibrium income here in the *IS–LM* model. The difference is explained by the crowding out of investment due to a higher interest rate.

Changes in Taxes

In the *IS–LM* model, changes in taxes affect the economy much the same as changes in government purchases do, except that taxes affect expenditure through consumption. Consider, for instance, a decrease in taxes of ΔT. The tax cut encourages consumers to spend more and, therefore, increases planned expenditure. The tax multiplier in the Keynesian cross tells us that this change in policy raises income at any given interest rate by $\Delta T \times MPC/(1 - MPC)$. Therefore, as [Figure 12-2](#) illustrates, the *IS* curve shifts to the right by this amount. The equilibrium of the economy moves from point A to point B. The tax cut raises both income and the interest rate. Once again, because the higher interest rate depresses investment, the increase in income is smaller in the *IS–LM* model than it is in the Keynesian cross.

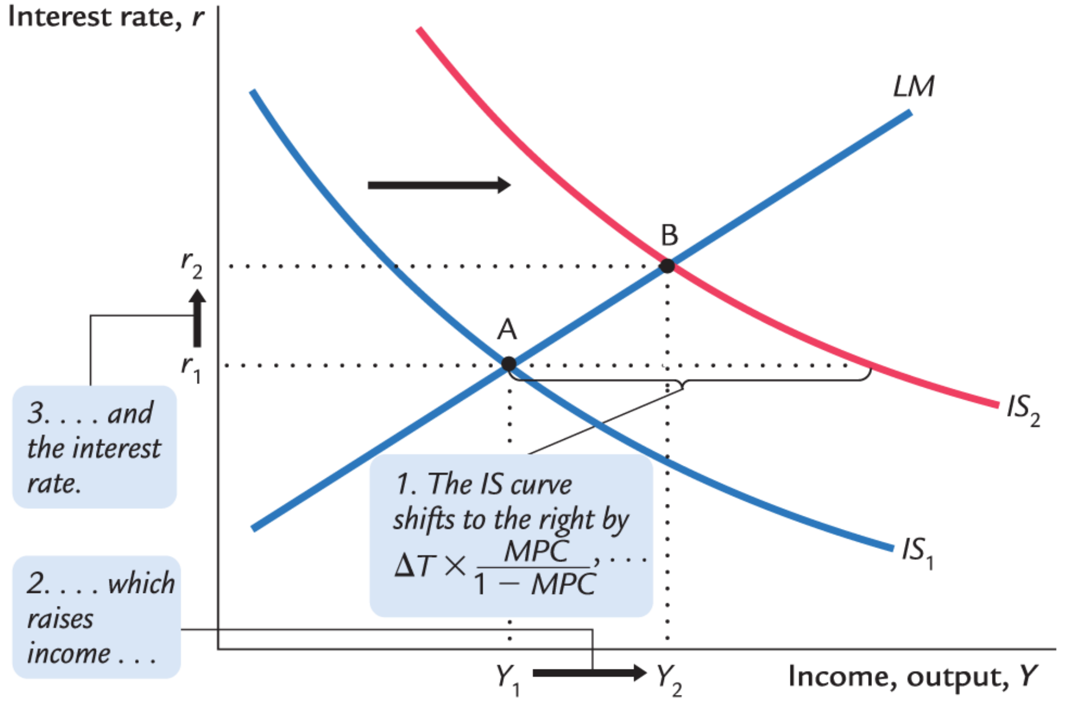


FIGURE 12-2 A Decrease in Taxes in the *IS–LM* Model A decrease in taxes shifts the *IS* curve to the right. The equilibrium moves from point A to point B. Income rises from Y_1 to Y_2, and the interest rate rises from r_1 to r_2.

How Monetary Policy Shifts the *LM* Curve and Changes the Short-Run Equilibrium

We now examine the effects of monetary policy. Recall that a change in the money supply alters the interest rate that equilibrates the money market for any given income and, thus, shifts the *LM* curve. The *IS–LM* model shows how a shift in the *LM* curve affects income and the interest rate.

Consider an increase in the money supply. An increase in *M* leads to an increase in real money balances M/P because the price level *P* is fixed in the short run. The theory of liquidity preference shows that for any given income, an increase in real money balances leads to a lower interest rate. Therefore, the *LM* curve shifts downward, as in Figure 12-3. The equilibrium moves from point A to point B. The increase in the money supply lowers the interest rate and raises income.

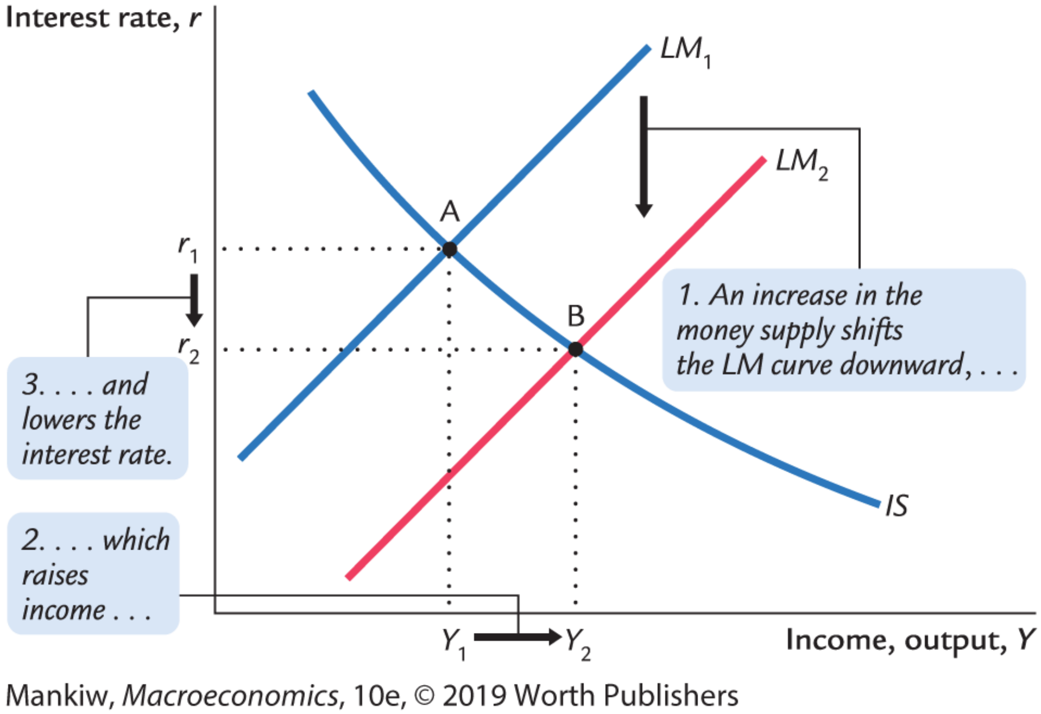


FIGURE 12-3 An Increase in the Money Supply in the *IS–LM* Model An increase in the money supply shifts the *LM* curve downward. The equilibrium moves from point A to point B. Income rises from Y1Y_1 to Y2,Y_2, and the interest rate falls from r1r_1 to r2.r_2.

Once again, to tell the story that explains the economy's adjustment from point A to point B, we rely on the building blocks of the *IS–LM* model—the Keynesian cross and the theory of liquidity preference. This time, we begin with the money market, where the monetary-policy action occurs. When the Fed increases the supply of money, people have more money than they want to hold at the prevailing interest rate. As a result, they start buying bonds or depositing this extra money in banks. The interest rate *r* then falls until people are willing to hold all the extra money that the Fed has created; this brings the money market to a new equilibrium. The lower interest rate, in turn, has ramifications for the goods market. A lower interest rate stimulates planned investment, which increases planned expenditure, production, and income *Y*.

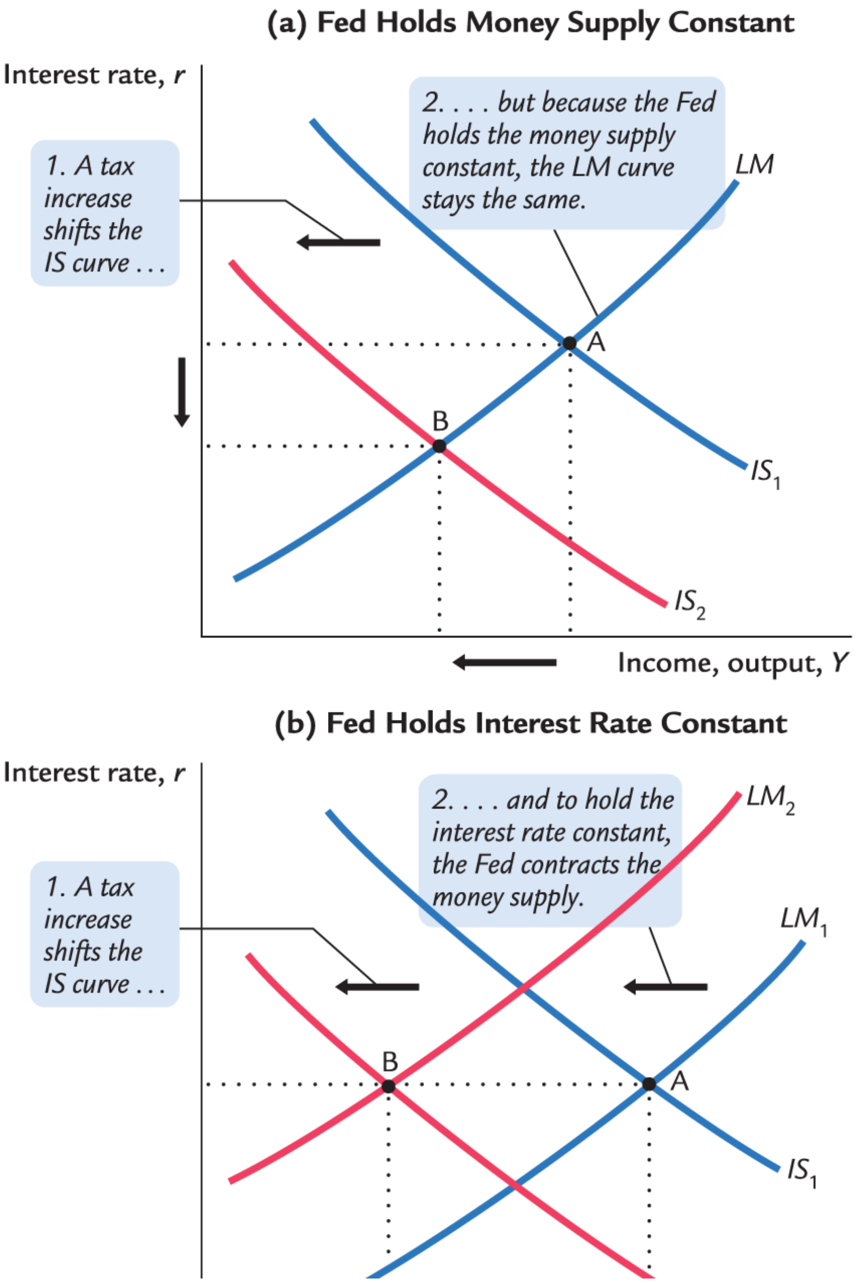
Thus, the *IS–LM* model shows that monetary policy influences income by changing the interest rate. This conclusion sheds light on our analysis of monetary policy in Chapter 10. In that chapter we showed that in the short run, when prices are sticky, an expansion in the money supply raises income. But we did not discuss *how* a monetary expansion induces greater spending on goods and services—a process called the **monetary transmission mechanism**. The *IS–LM* model shows an important part of that mechanism: *An increase in the money supply lowers the interest rate, which stimulates investment and thereby expands the demand for goods and services*. The next chapter shows that in open economies, the exchange rate also has a role in the monetary transmission mechanism; for large economies such as that of the United States, however, the interest rate has the leading role.

The Interaction Between Monetary and Fiscal Policy

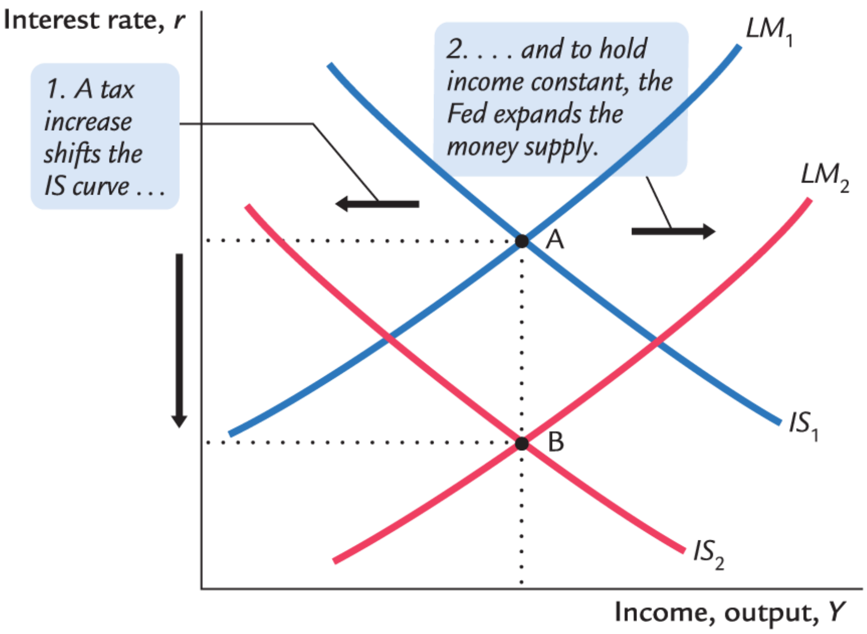
When analyzing any change in monetary or fiscal policy, it is important to keep in mind that the policymakers who control these policy tools are aware of what the other policymakers are doing. A change in one policy, therefore, may influence the other, and this interdependence may alter the impact of a policy change.

For example, suppose Congress raises taxes. What effect will this policy have on the economy? According to the *IS*–*LM* model, the answer depends on how the Fed responds to the tax increase.

Figure 12-4 shows three possible outcomes. In panel (a), the Fed holds the money supply constant. The tax increase shifts the *IS* curve to the left. Income falls (because higher taxes reduce consumer spending), and the interest rate falls (because lower income reduces the demand for money). The fall in income indicates that the tax hike causes a recession.



(c) Fed Holds Income Constant



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 12-4 The Response of the Economy to a Tax Increase How the economy responds to a tax increase depends on how the central bank responds. In panel (a) the Fed holds the money supply constant. In panel (b) the Fed holds the interest rate constant by reducing the money supply. In panel (c) the Fed holds income constant by increasing the money supply. In each case, the economy moves from point A to point B.

In panel (b), the Fed wants to hold the interest rate constant. In this case, when the tax increase shifts the *IS* curve to the left, the Fed must decrease the money supply to keep the interest rate at its original level. This fall in the money supply shifts the *LM* curve upward. The interest rate does not fall, but income falls by a larger amount than if the Fed had held the money supply constant. Whereas in panel (a) the lower interest rate stimulated investment and partially offset the contractionary effect of the tax hike, in panel (b) the Fed deepens the recession by keeping the interest rate high.

In panel (c), the Fed wants to prevent the tax increase from lowering income. It must, therefore, expand the money supply and shift the *LM* curve downward enough to offset the shift in the *IS* curve. In this case, the tax increase does not cause a recession, but it does cause a large fall in the interest rate. Although income does not change, the combination of a tax increase and a monetary expansion changes the allocation of the economy's resources. The higher taxes depress consumption, while the lower interest rate stimulates investment. Income is unchanged because these two effects exactly balance.

From this example we can see that the impact of fiscal policy depends on the policy the Fed pursues—that is, on whether it holds the money supply, the interest rate, or income constant. More generally, whenever analyzing a change in one policy, we must make an assumption about its effect on the other policy. The most appropriate assumption depends on the case at hand and the many political considerations that lie behind economic policymaking.

Shocks in the *IS–LM* Model

Calvin and Hobbes © 1992 Watterson Dist. by Andrews McMeel.

Because the *IS–LM* model shows how national income is determined in the short run, we can use the model to examine how various economic disturbances affect income. So far we have seen how changes in fiscal policy shift the *IS* curve and how changes in monetary policy shift the *LM* curve. Similarly, we can group other disturbances into two categories: shocks to the *IS* curve and shocks to the *LM* curve.

Shocks to the *IS* curve are exogenous changes in the demand for goods and services. Some economists, including Keynes, have emphasized that such changes in demand can arise from investors' *animal spirits*—exogenous and perhaps self-fulfilling waves of optimism and pessimism. For example, suppose firms become pessimistic about the future and react by building fewer new factories. This reduction in the demand for investment goods causes a contractionary shift in the investment function: at every interest rate, firms want to invest less. The fall in investment reduces planned expenditure and shifts the *IS* curve to the left, reducing income and employment. This fall in equilibrium income in part validates the firms' initial pessimism.

Shocks to the *IS* curve may also arise from changes in the demand for consumer goods. Suppose, for instance, the election of a popular president increases consumer confidence in the economy. This induces consumers to save less for the future and consume more today. We can interpret this change as an upward shift in the consumption function. This shift in the consumption function increases planned expenditure, shifts the *IS* curve to the right, and increases equilibrium income.

Shocks to the *LM* curve arise from exogenous changes in the demand for money. For example, suppose new restrictions on credit card availability increase the amount of money people want to hold. According to the theory of liquidity preference, when money demand rises, the interest rate necessary to equilibrate the money market is higher (for any given income and money supply). Hence, an increase in money demand shifts the *LM* curve upward, which tends to raise the interest rate and depress income.

In summary, several kinds of events can cause economic fluctuations by shifting the *IS* curve or the *LM* curve. Remember, however, that such fluctuations are not inevitable. Policymakers can try to use the tools of monetary and fiscal policy to offset exogenous shocks. If policymakers are sufficiently quick and skillful (admittedly, a big if), shocks to the *IS* or *LM* curves need not lead to fluctuations in income or employment.

CASE STUDY

The U.S. Recession of 2001

In 2001, the U.S. economy experienced a pronounced slowdown in economic activity. The unemployment rate rose from 3.9 percent in September 2000 to 4.9 percent in August 2001, and then to 6.3 percent in June 2003. In many ways, the slowdown looked like a typical recession driven by a fall in aggregate demand.

Three notable shocks explain this event. The first was a decline in the stock market. During the 1990s, the stock market experienced a boom of historic proportions, as investors became optimistic about the prospects of the new information technology. Some economists viewed the optimism as excessive at the time, and in hindsight this proved to be the case. When the optimism faded, average stock prices fell by about 25 percent from August 2000 to August 2001. The fall in the market reduced household wealth and thus consumer spending. In addition, the declining perceptions of the profitability of the new technologies led to a fall in investment spending. In the language of the *IS–LM* model, the *IS* curve shifted to the left.

The second shock was the terrorist attacks on New York City and Washington, DC, on September 11, 2001. In the week after the attacks, the stock market fell another 12 percent, which at the time was the biggest weekly loss since the Great Depression of the 1930s. Moreover, the attacks increased uncertainty about what the future would hold. Uncertainty can reduce spending because households and firms postpone some of their plans until the uncertainty is resolved. Thus, the terrorist attacks shifted the *IS* curve farther to the left.

The third shock was a series of accounting scandals at some of the nation's most prominent corporations, including Enron and WorldCom. These scandals resulted in the bankruptcy of some companies that had fraudulently represented themselves as more profitable than they were, criminal convictions for the executives responsible for the fraud, and new laws aimed at regulating corporate accounting standards more thoroughly. These events further depressed stock prices and discouraged business investment—a third leftward shift in the *IS* curve.

Fiscal and monetary policymakers responded quickly to these events. Congress passed a major tax cut in 2001, including an immediate tax rebate, and a second major tax cut in 2003. One goal of these tax cuts was to stimulate consumer spending. (See the Case Study "Cutting Taxes to Stimulate the Economy: The Kennedy and Bush Tax Cuts" in Chapter 11.) In addition, after the 2001 terrorist attacks, Congress increased government spending by appropriating funds to assist in New York's recovery and to bail out the ailing airline industry. These fiscal measures shifted the *IS* curve to the right.

At the same time, the Fed pursued expansionary monetary policy, shifting the *LM* curve to the right. Money growth accelerated, and interest rates fell. The interest rate on three-month Treasury bills fell from 6.2 percent in November 2000 to 3.4 percent in August 2001, just before the terrorist attacks. After the attacks and corporate scandals hit the economy, the Fed increased its monetary stimulus, and the Treasury bill rate fell to 0.9 percent in July 2003—the lowest level in many decades.

Expansionary monetary and fiscal policy had the intended effects. Economic growth picked up in the second half of 2003 and was strong throughout 2004. By July 2005, the unemployment rate was back down to 5.0 percent, and it stayed at or below that level for the next several years. Unemployment would begin rising again in 2008, however, when the economy experienced another recession. The causes of the 2008 recession are examined in another Case Study later in this chapter. ■

What Is the Fed's Policy Instrument—The Money Supply or the Interest Rate?

Our analysis of monetary policy has been based on the assumption that the Fed influences the economy by controlling the money supply. By contrast, when the media report on changes in Fed policy, they often just say that the Fed has raised or lowered interest rates. Which is right? Even though these two views may seem different, both are correct, and it is important to understand why.

In recent years, the Fed has used the *federal funds rate*—the interest rate that banks charge one another for overnight loans—as its short-term policy instrument. When the Federal Open Market Committee (FOMC) meets about every six weeks to set monetary policy, it votes on a target for this interest rate that will apply until the next meeting. After the meeting is over, the Fed's bond traders (who are in New York) are told to conduct the open-market operations necessary to hit that target. These open-market operations change the money supply and shift the *LM* curve so that the equilibrium interest rate (determined by the intersection of the *IS* and *LM* curves) equals the target interest rate that the FOMC has chosen.

As a result of this operating procedure, monetary policy is often discussed in terms of changing interest rates. Remember, however, that behind these changes in interest rates are the necessary changes in the money supply. A newspaper might report that "the Fed has lowered interest rates." This statement should be interpreted as follows: "the FOMC has instructed the Fed bond traders to buy bonds in open-market operations so as to increase the money supply, shift the *LM* curve, and reduce the equilibrium interest rate to hit a new lower target."

Why has the Fed chosen to use an interest rate, rather than the money supply, as its short-term policy instrument? One possible answer is that shocks to the *LM* curve are more prevalent than shocks to the *IS* curve. When the Fed targets interest rates, it automatically offsets *LM* shocks by adjusting the money supply, although this policy exacerbates *IS* shocks. If *LM* shocks are the more prevalent type, then a policy of targeting the interest rate leads to greater economic stability than a policy of targeting the money supply. (Problem 8 at the end of this chapter asks you to analyze this issue more fully.)

In Chapter 15 we extend our theory of short-run fluctuations to explicitly include a monetary policy that targets the interest rate and changes its target in response to economic conditions. The *IS*–*LM* model presented here is a useful foundation for that more complicated and realistic analysis. One lesson from the *IS–LM* model is that when a central bank sets the money supply, it determines the equilibrium interest rate. Thus, in some ways, setting the money supply and setting the interest rate are two sides of the same coin.

12-2 *IS–LM* as a Theory of Aggregate Demand

We have been using the *IS–LM* model to explain national income in the short run when the price level is fixed. To see how the *IS–LM* model fits into the model of aggregate supply and aggregate demand introduced in Chapter 10, we now examine what happens in the *IS–LM* model when the price level changes. By examining the effects of changing the price level, we can deliver what was promised when we began our study of the *IS–LM* model: a theory to explain the position and slope of the aggregate demand curve.

From the *IS–LM* Model to the Aggregate Demand Curve

Recall from Chapter 10 that the aggregate demand curve describes a relationship between the price level and national income. In Chapter 10 this relationship was derived from the quantity theory of money. That analysis showed that for a given money supply, a higher price level implies lower income. Increases in the money supply shift the aggregate demand curve to the right, and decreases in the money supply shift the aggregate demand curve to the left.

To understand the determinants of aggregate demand more fully, we now use the *IS–LM* model, rather than the quantity theory, to derive the aggregate demand curve. First, we use the *IS–LM* model to show why national income falls as the price level rises—that is, why the aggregate demand curve is downward sloping. Second, we examine what causes the aggregate demand curve to shift.

To explain why the aggregate demand curve slopes downward, we examine what happens in the *IS–LM* model when the price level changes. This is done in Figure 12-5. For any given money supply *M*, a higher price level *P* reduces the supply of real money balances M/P. A lower supply of real money balances shifts the *LM* curve upward, which raises the equilibrium interest rate and lowers equilibrium income, as shown in panel (a). Here the price level rises from P_1 to P_2, and income falls from Y_1 to Y_2. The aggregate demand curve in panel (b) plots this negative relationship between national income and the price level. In other words, the aggregate demand curve shows the set of equilibrium points that arise in the *IS–LM* model as we vary the price level and see what happens to income.

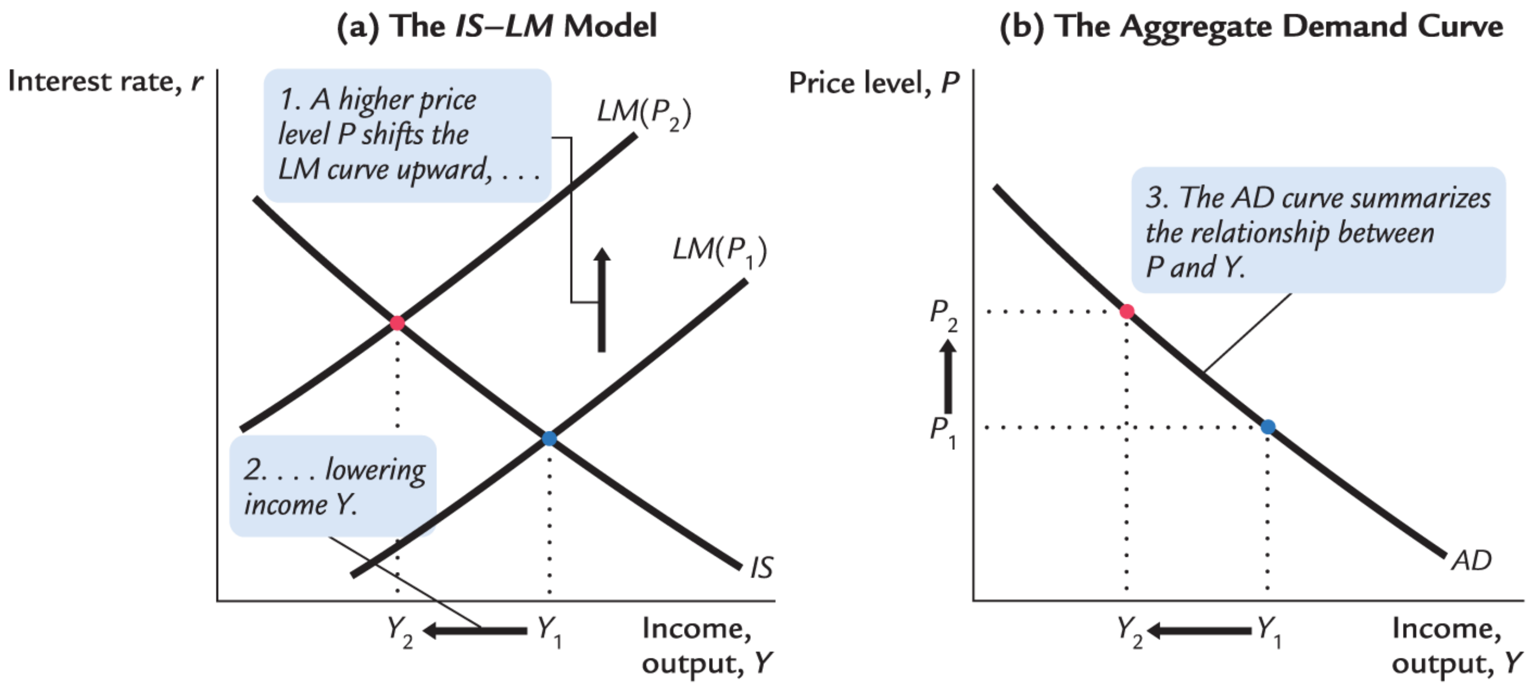


FIGURE 12-5 Deriving the Aggregate Demand Curve with the *IS–LM* Model Panel (a) shows the *IS–LM* model: an increase in the price level from P1P_1 to P2P_2 lowers real money balances and thus shifts the *LM* curve upward. The shift in the *LM* curve lowers income from Y1Y_1 to Y2.Y_2. Panel (b) shows the aggregate demand curve summarizing this relationship between the price level and income: the higher the price level, the lower the level of income.

What causes the aggregate demand curve to shift? Because the aggregate demand curve summarizes the results from the *IS–LM* model, events that shift the *IS* curve or the *LM* curve (for a given price level) cause the aggregate demand curve to shift. For instance, an increase in the money supply raises income in the *IS–LM* model for any given price level, shifting the aggregate demand curve to the right, as shown in panel (a) of Figure 12-6. Similarly, an increase in government purchases or a decrease in taxes raises income in the *IS–LM* model for a given price level, also shifting the aggregate demand curve to the right, as shown in panel (b). Conversely, a decrease in the money supply, a decrease in government purchases, or an increase in taxes lowers income in the *IS–LM* model and shifts the aggregate demand curve to the left. Anything that changes income in the *IS–LM* model other than a change in the price level causes a shift in the aggregate demand curve. The factors shifting aggregate demand include not only monetary and fiscal policy but also shocks to the goods market (the *IS* curve) and shocks to the money market (the *LM* curve).

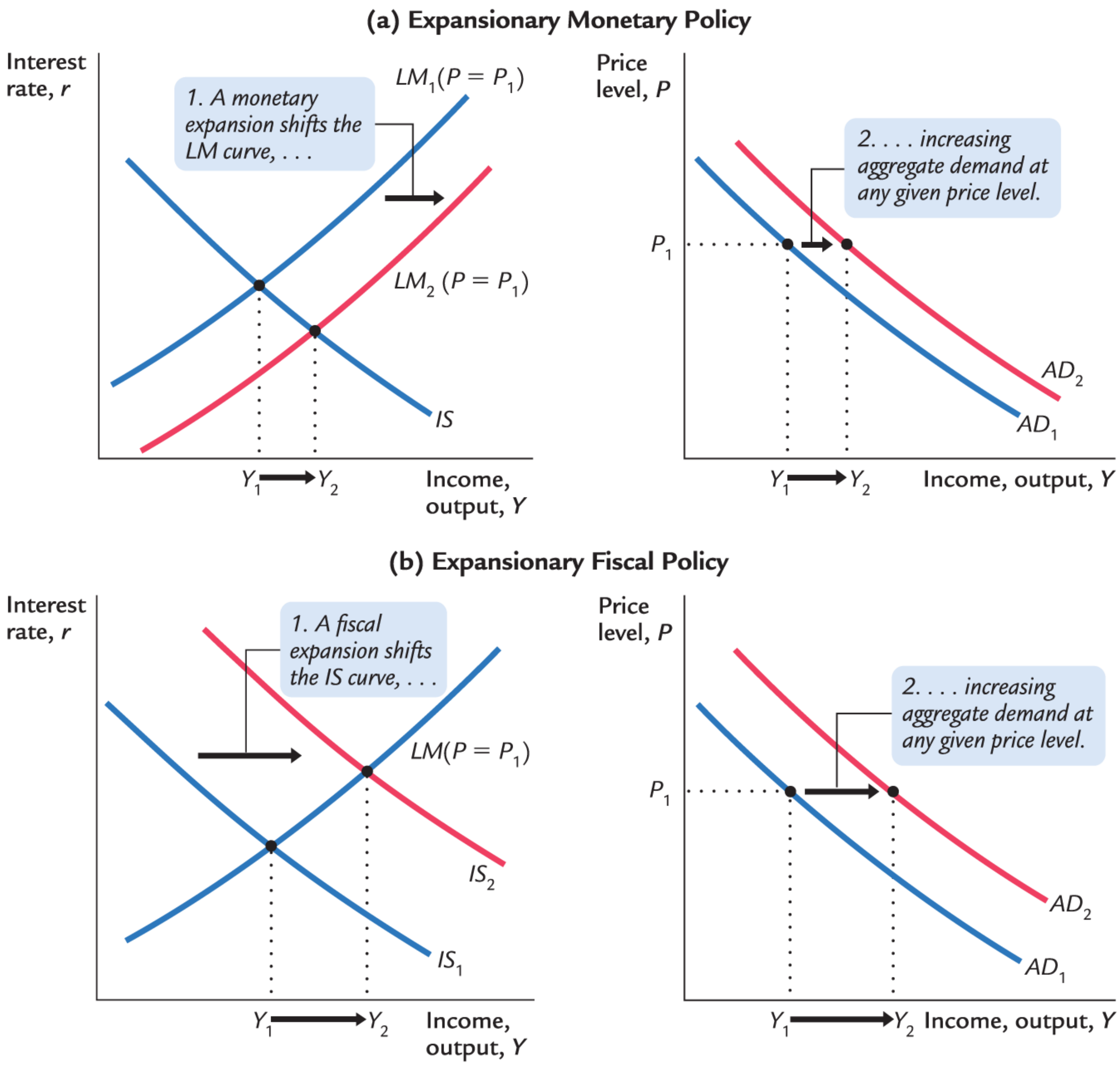


FIGURE 12-6 How Monetary and Fiscal Policies Shift the Aggregate Demand Curve Panel (a) shows a monetary expansion. For any given price level, an increase in the money supply raises real money balances, shifts the *LM* curve downward, and raises income. Hence, an increase in the money supply shifts the aggregate demand curve to the right. Panel (b) shows a fiscal expansion, such as an increase in government purchases or a decrease in taxes. The fiscal expansion shifts the *IS* curve to the right and, for any given price level, raises income. Hence, a fiscal expansion shifts the aggregate demand curve to the right.

We can summarize these results as follows: *a change in income in the* IS–LM *model resulting from a change in the price level represents a movement along the aggregate demand curve. A change in income in the* IS–LM *model for a given price level represents a shift in the aggregate demand curve.*

The *IS–LM* Model in the Short Run and Long

Run

The *IS–LM* model is designed to explain the economy in the short run when the price level is fixed. Yet, now that we have seen how a change in the price level influences the equilibrium in the *IS–LM* model, we can also use the model to describe the economy in the long run when the price level adjusts to ensure that the economy produces at its natural rate. By using the *IS–LM* model to describe the long run, we can show clearly how the Keynesian model of income determination differs from the classical model of Chapter 3.

Panel (a) of Figure 12-7 shows the three curves that are necessary for understanding the short-run and long-run equilibria: the *IS* curve, the *LM* curve, and the vertical line representing the natural level of output $\overline{Y}$. The *LM* curve is, as always, drawn for a fixed price level P_1. The short-run equilibrium of the economy is point K, where the *IS* curve crosses the *LM* curve. Notice that in this short-run equilibrium, the economy's income is less than its natural level.

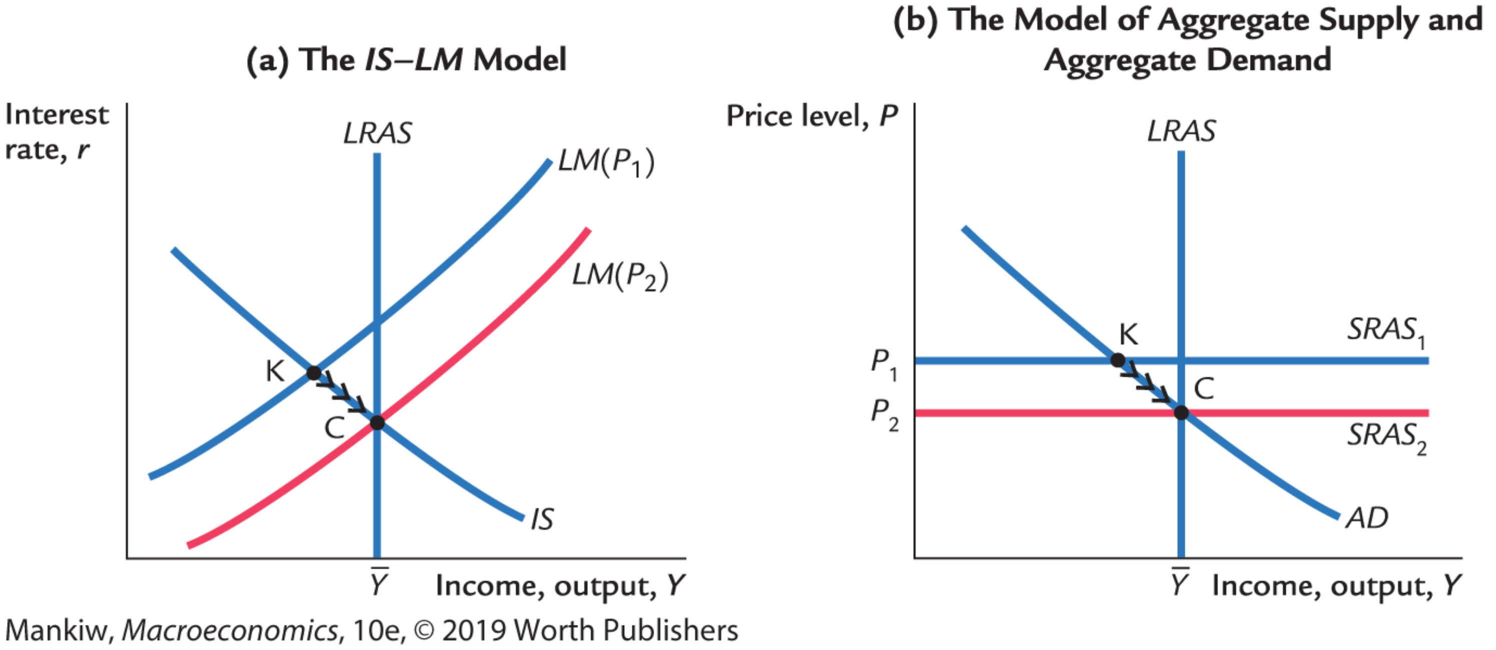


FIGURE 12-7 The Short-Run and Long-Run Equilibria We can compare the short-run and long-run equilibria using either the *IS–LM* diagram in panel (a) or the aggregate supply–aggregate demand diagram in panel (b). In the short run, the price level is stuck at P_1. The short-run equilibrium of the economy is therefore point K. In the long run, the price level adjusts so that the economy is at the natural level of output. The long-run equilibrium is therefore point C.

Panel (b) of Figure 12-7 shows the same situation in the diagram of aggregate supply and aggregate demand. At the price level P_1, the quantity of output demanded is below the natural level. In other words, at the existing price level, there is insufficient demand for goods and services to keep the economy producing at its potential.

In these two diagrams we can examine the short-run equilibrium at which the economy finds itself and the long-run equilibrium toward which the economy gravitates. Point K describes the short-run equilibrium because it assumes that the price level is stuck at P_1. Eventually, the low demand for goods and services causes prices to fall, and the economy moves back toward its natural rate. When the price level reaches P_2, the economy is at point C, the long-run equilibrium. The diagram of aggregate supply and aggregate demand shows that at point C, the quantity of goods and services demanded equals the natural level of output. This long-run equilibrium is achieved in the *IS–LM* diagram by a shift in the *LM* curve: the fall in the price level raises real money balances and therefore shifts the *LM* curve to the right.

We can now see the key difference between the Keynesian and classical approaches to the determination of national income. The Keynesian assumption (represented by point K) is that prices are stuck. Depending on monetary policy, fiscal policy, and the other determinants of aggregate demand, output may deviate from its natural level. The classical assumption (represented by point C) is that prices are flexible. The price level adjusts to ensure that national income is always at its natural level.

To make the same point somewhat differently, we can think of the economy as being described by three equations. The first two are the *IS* and *LM* equations:

Y=C(Y−T)+I(r)+GIS,M/P=L(r, Y)LM.

$$\begin{aligned} Y &= C(Y-T)+I(r)+G \quad && IS, \\ M/P &= L(r,\, Y) && LM. \end{aligned}$$

The *IS* equation describes the equilibrium in the goods market, and the *LM* equation describes the equilibrium in the money market. These *two* equations contain *three* endogenous variables: *Y*, *P*, and *r*. To complete the system, we need a third equation. The Keynesian approach completes the model with the assumption of fixed prices, so the Keynesian third equation is

P=P1.

$$P = P_1.$$

This assumption implies that the remaining two variables *r* and *Y* must adjust to satisfy the remaining two equations *IS* and *LM*. The classical approach completes the model with the assumption that output reaches its natural level, so the classical third equation is

Y=Y¯.

$$Y = \overline{Y}.$$

This assumption implies that the remaining two variables *r* and *P* must adjust to satisfy the remaining two equations *IS* and *LM*. Thus, the classical approach fixes output and allows the price level to adjust to satisfy the goods and money market equilibrium conditions, whereas the Keynesian approach fixes the price level and lets output move to satisfy the equilibrium conditions.

Which assumption is most appropriate? The answer depends on the time horizon. The classical assumption best describes the long run. Hence, our long-run analyses of national income in Chapter 3 and prices in Chapter 5 assume that output equals its natural level. The Keynesian assumption best describes the short run. Therefore, our analysis of short-run fluctuations assumes a fixed price level.

12-3 The Great Depression

Now that we have developed the model of aggregate demand, let's use it to address the question that motivated Keynes: what caused the Great Depression? Even today, almost a century after the event, economists debate the causes of this major downturn. The Great Depression provides an extended case study to show how economists use the *IS*–*LM* model to analyze economic fluctuations.[1]

Before turning to the explanations economists have proposed, look at Table 12-1, which presents some statistics regarding the Depression. These statistics are the battlefield on which debates about the Depression takes place. What do you think happened? An *IS* shift? An *LM* shift? Or something else?

TABLE 12-1 What Happened During the Great Depression?

Year	Unemployment Rate (1)	Real GNP (2)	Consumption (2)	Investment (2)	Government Purchases (2)	Nominal Interest Rate (3)	Money Supply (4)	Price Level (5)	Inflation (6)	Real Money Balances (7)
1929	3.2	203.6	139.6	40.4	22.0	5.9	26.6	50.6	—	52.6
1930	8.9	183.5	130.4	27.4	24.3	3.6	25.8	49.3	−2.6	52.3
1931	16.3	169.5	126.1	16.8	25.4	2.6	24.1	44.8	−10.1	54.5
1932	24.1	144.2	114.8	4.7	24.2	2.7	21.1	40.2	−9.3	52.5
1933	25.2	141.5	112.8	5.3	23.3	1.7	19.9	39.3	−2.2	50.7
1934	22.0	154.3	118.1	9.4	26.6	1.0	21.9	42.2	7.4	51.8
1935	20.3	169.5	125.5	18.0	27.0	0.8	25.9	42.6	0.9	60.8
1936	17.0	193.2	138.4	24.0	31.8	0.8	29.6	42.7	0.2	62.9
1937	14.3	203.2	143.1	29.9	30.8	0.9	30.9	44.5	4.2	69.5
1938	19.1	192.9	140.2	17.0	33.9	0.8	30.5	43.9	−1.3	69.5
1939	17.2	209.4	148.2	24.7	35.2	0.6	34.2	43.2	−1.6	79.1
1940	14.6	227.2	155.7	33.0	36.4	0.6	39.7	43.9	1.6	90.3

Data from: Historical Statistics of the United States, Colonial Times to 1970, Parts I and II (Washington, DC: U.S. Department of Commerce, Bureau of Census, 1975).

Note: (1) The unemployment rate is series D9. (2) Real GNP, consumption, investment, and government purchases are series F3, F48, F52, and F66, and are measured in billions of 1958 dollars. (3) The interest rate is the prime Commercial Paper rate, 4–6 months, series X445. (4) The money supply is series X414, currency plus demand deposits, measured in billions of dollars. (5) The price level is the GNP deflator (1958=100), series E1. (6) The inflation rate is the percentage change in the price level series. (7) Real money balances, calculated by dividing the money supply by the price level and multiplying by 100, are in billions of 1958 dollars.

The Spending Hypothesis: Shocks to the *IS* Curve

Table 12-1 shows that the decline in income in the early 1930s coincided with falling interest rates. This fact has led some economists to suggest that the cause of the decline may have been a contractionary shift in the *IS* curve. This view is sometimes called the *spending hypothesis* because it places primary blame for the Depression on an exogenous fall in spending on goods and services.

Economists have proposed several explanations for this decline in spending. Some argue that a downward shift in the consumption function caused the contractionary shift in the *IS* curve. The stock market crash of 1929 may have been partly responsible for this shift: by reducing wealth and increasing uncertainty about the prospects of the U.S. economy, the crash may have induced consumers to save more of their income rather than spend it.

Others explain the decline in spending by pointing to the large drop in investment in housing. Some economists believe that the residential investment boom of the 1920s was excessive and that once this "overbuilding" became apparent, the demand for residential investment declined drastically. Another possible explanation for the fall in residential investment is the reduction in immigration in the 1930s: a more slowly growing population demands less new housing.

Once the Depression began, several events occurred that could have reduced spending further. One is the failure of many banks in the early 1930s, in part because of inadequate bank regulation and in part because of the Fed's reluctance to play an active role as lender of last resort when runs on the banks began. As we discuss more fully in Chapter 18, banks play the crucial role of getting the funds available for investment to those households and firms that can best use them. The closing of many banks in the early 1930s may have prevented some businesses from getting the funds they needed for capital investment and, therefore, may have led to a further contraction in investment spending.[2]

The fiscal policy of the 1930s also contributed to the contractionary shift in the *IS* curve. Politicians at that time were more concerned with balancing the budget than with using fiscal policy to keep production and employment at their natural levels. The Revenue Act of 1932 increased various taxes, especially those affecting lower- and middle-income consumers.[3] The Democratic platform that year expressed concern about the budget deficit and advocated an "immediate and drastic reduction of governmental expenditures." In the midst of historically high unemployment, policymakers searched for ways to raise taxes and reduce government spending.

There are, therefore, several ways to explain a contractionary shift in the *IS* curve. Keep in mind that these different views may all be true. Rather than having a single explanation, the massive decline in spending may be the result of many contractionary forces hitting the economy at the same time.

The Money Hypothesis: A Shock to the *LM* Curve

Table 12-1 shows that the money supply fell 25 percent from 1929 to 1933, during which time the unemployment rate rose from 3.2 percent to 25.2 percent. This fact provides the motivation and support for what is called the *money hypothesis*, which places primary blame for the Depression on the Fed for allowing the money supply to fall by such a large amount.[4] The best-known advocates of this interpretation are Milton Friedman and Anna Schwartz, who defended it in their treatise on U.S. monetary history. Friedman and Schwartz argue that contractions in the money supply have caused most economic downturns and that the Great Depression is a dramatic example.

Using the *IS*–*LM* model, we might interpret the money hypothesis as explaining the Depression by a contractionary shift in the *LM* curve. Seen in this way, however, the money hypothesis runs into two problems.

The first problem is the behavior of *real* money balances. Monetary policy leads to a contractionary shift in the *LM* curve only if real money balances fall. Yet from 1929 to 1931 real money balances rose slightly because the fall in the money supply was accompanied by an even greater fall in the price level. Although the monetary contraction may have been responsible for the rise in unemployment from 1931 to 1933, when real money balances did fall, it cannot easily explain the initial downturn from 1929 to 1931.

The second problem for the money hypothesis is the behavior of interest rates. If a contractionary shift in the *LM* curve triggered the Depression, we should have observed higher interest rates. Yet nominal interest rates fell continuously from 1929 to 1933.

These two reasons appear sufficient to reject the view that the Depression was instigated by a contractionary shift in the *LM* curve. But was the fall in the money stock irrelevant? Next, we turn to another mechanism through which monetary policy might have been responsible for the severity of the Depression—the deflation of the 1930s.

The Money Hypothesis Again: The Effects of Falling Prices

From 1929 to 1933 the price level fell 22 percent. Many economists blame this deflation for the Great Depression's severity. They argue that the deflation may have turned what in 1931 was a typical downturn into an unprecedented period of high unemployment and depressed income. If correct, this argument gives new life to the money hypothesis. Because the falling money supply was, plausibly, responsible for the falling price level, it could have been responsible for the severity of the Depression. To evaluate this argument, we must discuss how changes in the price level affect income in the *IS*–*LM* model.

The Stabilizing Effects of Deflation

In the *IS*–*LM* model we have developed so far, falling prices raise income. For any given supply of money *M*, a lower price level implies higher real money balances M/PM/P. An increase in real money balances causes an expansionary shift in the *LM* curve, which leads to higher income.

Another channel through which falling prices expand income is called the **Pigou effect**. Arthur Pigou, a prominent economist in the 1930s, pointed out that real money balances are part of households' wealth. As prices fall and real money balances rise, consumers should feel wealthier and spend more. This increase in consumer spending should cause an expansionary shift in the *IS* curve, also leading to higher income.

These two reasons led some economists in the 1930s to believe that falling prices would help stabilize the economy. That is, they thought that a decline in the price level would push the economy back toward full employment. However, other economists were less confident in the economy's ability to correct itself. They pointed to other effects of falling prices, to which we now turn.

The Destabilizing Effects of Deflation

Economists have proposed two theories to explain how falling prices could depress income rather than raise it. The first, called the **debt-deflation theory**, describes the effects of unexpected falls in the price level. The second explains the effects of expected deflation.

The debt-deflation theory begins with an observation from Chapter 5: unanticipated changes in the price level redistribute wealth between debtors and creditors. If a debtor owes a creditor $1,000, then the real value of the debt is $1,000/P,$\$1,000/P,$ where *P* is the price level. A fall in the price level raises the debt's real value; the debtor must repay the creditor a larger amount of purchasing power. Thus, an unexpected deflation enriches creditors and impoverishes debtors.

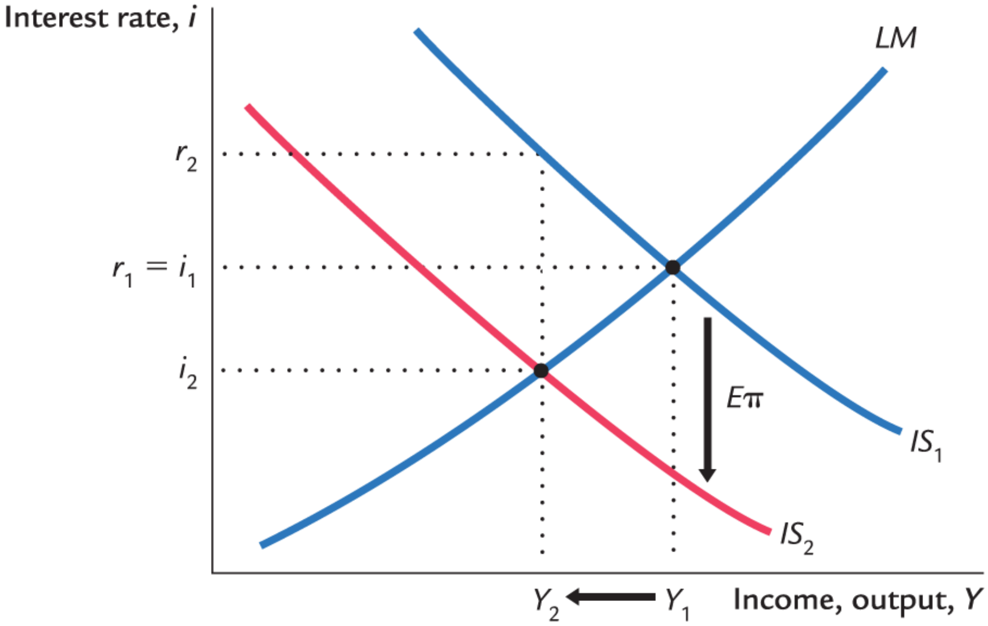
The debt-deflation theory then posits that this redistribution of wealth affects spending on goods and services. In response to the redistribution from debtors to creditors, debtors spend less and creditors spend more. If these two groups have equal spending propensities, there is no aggregate impact. But debtors may have higher propensities to spend than creditors; perhaps that is why the debtors are in debt in the first place. In this case, debtors reduce their spending by more than creditors raise theirs. The net effect is a reduction in overall spending, leading to a contractionary shift in the *IS* curve and lower national income.

To understand how *expected* changes in prices can affect income, we need to add a new variable to the *IS–LM* model. Our discussion of the model so far has not distinguished between the nominal and real interest rates. Yet we know from previous chapters that investment depends on the real interest rate and that money demand depends on the nominal interest rate. If *i* is the nominal interest rate and $E\pi$ is expected inflation, then the *ex ante* real interest rate is $i - E\pi$. We can now write the *IS–LM* model as

$$\begin{aligned} Y &= C(Y-T) + I(i - E\pi) + G \qquad IS, \\ M/P &= L(i, Y) \qquad LM. \end{aligned}$$

Expected inflation enters as a variable in the *IS* curve. Thus, changes in expected inflation shift the *IS* curve.

Let's use this extended *IS–LM* model to examine how changes in expected inflation affect income. We begin by assuming that everyone expects the price level to remain the same. In this case, there is no expected inflation $(E\pi = 0)$, and these two equations produce the familiar *IS–LM* model. Figure 12-8 depicts this initial situation; the intersection of the *LM* curve and the *IS* curve labeled IS_1 determines the nominal and real interest rates, which for now are the same.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 12-8 Expected Deflation in the *IS–LM* Model An expected deflation (a negative value of $E\pi$) raises the real interest rate for any given nominal interest rate, and this depresses investment spending. The reduction in investment shifts the *IS* curve downward. Income falls from Y_1 to Y_2. The nominal interest rate falls from i_1 to i_2, and the real interest rate rises from r_1 to r_2.

Now suppose everyone suddenly expects the price level to fall in the future, making $E\pi$ negative. The real interest rate is now higher at any given nominal interest rate. This increase in the real interest rate depresses planned investment spending, shifting the *IS* curve from IS_1 to IS_2. (The vertical distance of the downward shift equals the expected deflation.) Thus, an expected deflation leads to a reduction in national income from Y_1 to Y_2. The nominal interest rate falls from i_1 to i_2, while the real interest rate rises from r_1 to r_2.

Here is the story behind this figure. When firms come to expect deflation, they become reluctant to borrow to buy investment goods because they believe they will have to repay these loans later in more valuable dollars. The fall in investment depresses planned expenditure, which in turn depresses income. The fall in income reduces the demand for money, and this reduces the nominal interest rate that equilibrates the money market. The nominal interest rate falls by less than the expected deflation, so the real interest rate rises.

Note that there is a common thread in these two stories of destabilizing deflation. In both, falling prices depress national income by causing a contractionary shift in the *IS* curve. Because a deflation of the size observed from 1929 to 1933 is unlikely except in the presence of a major contraction in the money supply, these two explanations assign some of the responsibility for the Depression—especially its severity—to the Fed. In other words, if falling prices are destabilizing, then a contraction in the money supply can lead to a fall in income, even without a decrease in real money balances or a rise in nominal interest rates.

Could the Depression Happen Again?

Economists study the Depression both to understand a major historic event and to help policymakers ensure that it will not happen again. To state with confidence whether this event could recur, we would need to know why it happened. Because there is not yet agreement on the causes of the Great Depression, it is impossible to rule out with certainty another depression of such magnitude.

Yet most economists believe that the mistakes that led to the Great Depression probably won't be repeated. The Fed seems unlikely to allow the money supply to fall by one-fourth. Many economists believe that the deflation of the early 1930s was responsible for the Depression's severity. And it seems likely that such a prolonged deflation was possible only in the presence of a falling money supply.

The fiscal-policy mistakes of the Depression are also unlikely to be repeated. Fiscal policy in the 1930s not only failed to help but actually further depressed aggregate demand. Few economists today would support rigid adherence to a balanced budget in the face of massive unemployment.

In addition, many modern institutions would help prevent the events of the 1930s from recurring. The system of federal deposit insurance makes widespread bank failures less likely. The income tax causes an automatic reduction in taxes when income falls, which stabilizes the economy. Finally, economists know more today than they did in the 1930s. Our knowledge of how the economy works, limited as it still is, should help policymakers formulate better policies to combat widespread unemployment.

CASE STUDY

The Financial Crisis and Great Recession of 2008 and 2009

In 2008 the U.S. economy experienced a financial crisis, followed by a deep economic downturn. Several of the developments during this time were reminiscent of events during the 1930s, causing many observers to fear that the economy might experience a second Great Depression.

The story of the 2008 crisis begins a few years earlier, with a substantial boom in the housing market. The boom had several sources. In part, it was fueled by low interest rates. As we saw in a previous Case Study in this chapter, the Fed lowered interest rates to historically low levels in the aftermath of the recession of 2001. Low interest rates helped the economy recover but also contributed to a rise in house prices by making it less expensive to get a mortgage and buy a home.

In addition, developments in the mortgage market made it easier for *subprime borrowers*—borrowers with higher risk of default based on their income and credit history—to get mortgages to buy homes. One of these developments was *securitization*, the process by which one mortgage originator makes loans and then sells them to an investment bank, which in turn bundles them together into a variety of "mortgage-backed securities" and then sells them to a third financial institution (such as a bank, a pension fund, or an insurance company). These securities pay a return as long as homeowners continue to repay their loans, but they lose value if homeowners default. Unfortunately, the ultimate holders of these mortgage-backed securities sometimes failed to fully appreciate the risks they were taking. Some economists blame insufficient regulation for these high-risk loans. Others believe the problem was not too little regulation but the wrong kind: some government policies encouraged this high-risk lending to make homeownership more attainable for low-income families.

Together, these forces drove up housing demand and prices. From 1995 to 2006, average house prices in the United States more than doubled. Some observers view this rise in house prices as a speculative bubble, as more people bought homes hoping and expecting that the prices would continue to rise.

The high price of housing, however, proved unsustainable. From 2006 to 2009, house prices nationwide fell about 30 percent. Such price fluctuations should not necessarily be a problem in a market economy. After all, price movements are needed to equilibrate supply and demand. But, in this case, the price decline had several problematic repercussions.

The first repercussion was a substantial rise in mortgage defaults and home foreclosures. During the housing boom, many homeowners had bought their homes with borrowed money and minimal down payments. When house prices declined, these homeowners were *underwater*: they owed more on their mortgages than their homes were worth. As a result, many of these homeowners stopped paying their loans. The banks servicing the mortgages responded to the defaults by taking the houses away in foreclosure procedures and then selling them off. The banks' goal was to recoup whatever they could. The increase in the number of homes for sale, however, exacerbated the downward spiral of house prices.

A second repercussion was large losses at the various financial institutions that owned mortgage-backed securities. By borrowing large sums to buy high-risk mortgages, these companies had bet that house prices would keep rising; when this bet turned bad, they found themselves at or near the point of bankruptcy. Even healthy banks stopped trusting one another and avoided interbank lending because it was hard to discern which institution would be the next to go out of business. Because of these large losses at financial institutions and the widespread fear and distrust, the ability of the financial system to make loans even to creditworthy customers was impaired. Chapter 18 discusses financial crises, including this one, in more detail.

A third repercussion was a substantial rise in stock market volatility. Many companies rely on the financial system to get the resources they need for business expansion or to help them manage their short-term cash flows. With the financial system less able to perform its normal operations, the profitability of many companies was called into question. Because it was hard to know how bad things would get, stock market volatility reached levels not seen since the 1930s.

Falling house prices, increasing foreclosures, financial instability, and higher volatility together led to a fourth repercussion: a decline in consumer confidence. In the midst of all the uncertainty, households started putting off spending plans. In particular, expenditure on durable goods such as cars and household appliances plummeted.

As a result of these events, the economy experienced a large contractionary shift in the *IS* curve. Production, income, and employment declined. The unemployment rate rose from 4.7 percent in October 2007 to 10.0 percent in October 2009.

Policymakers responded vigorously as the crisis unfolded. First, the Fed cut its target for the federal funds rate from 5.25 percent in September 2007 to about zero in December 2008. Second, in October 2008, Congress appropriated $700 billion for the Treasury to use to rescue the financial system. In large part these funds were used for equity injections into banks. That is, the Treasury put funds into the banking system, which the banks could then use to make loans; in exchange for these funds, the U.S. government temporarily became a part owner of these banks. Third, as discussed in Chapter 11, one of Barack Obama’s first acts as president was to support a major increase in government spending to expand aggregate demand. Finally, the Fed engaged in various unconventional monetary policies, such as buying long-term bonds, to lower long-term interest rates and thereby encourage borrowing and private spending.

In the end, policymakers can take some credit for having averted another Great Depression. Unemployment rose to only 10 percent, compared with 25 percent in 1933. Other data tell a similar story. Figure 12-9 compares the path of industrial production during the Great Depression of the 1930s and during the Great Recession of 2008–2009. (Industrial production measures the output of the nation’s manufacturers, mines, and utilities. Because of the consistency of its data sources, it is one of the more reliable time series for historical comparisons of short-run fluctuations.) The figure shows that, in the Great Depression, industrial production declined for about three years, falling by more than 50 percent, and it took more than seven years to return to its previous peak. By contrast, in the Great Recession, industrial production declined for only a year and half, fell only 17 percent, and took less than six years to recover.

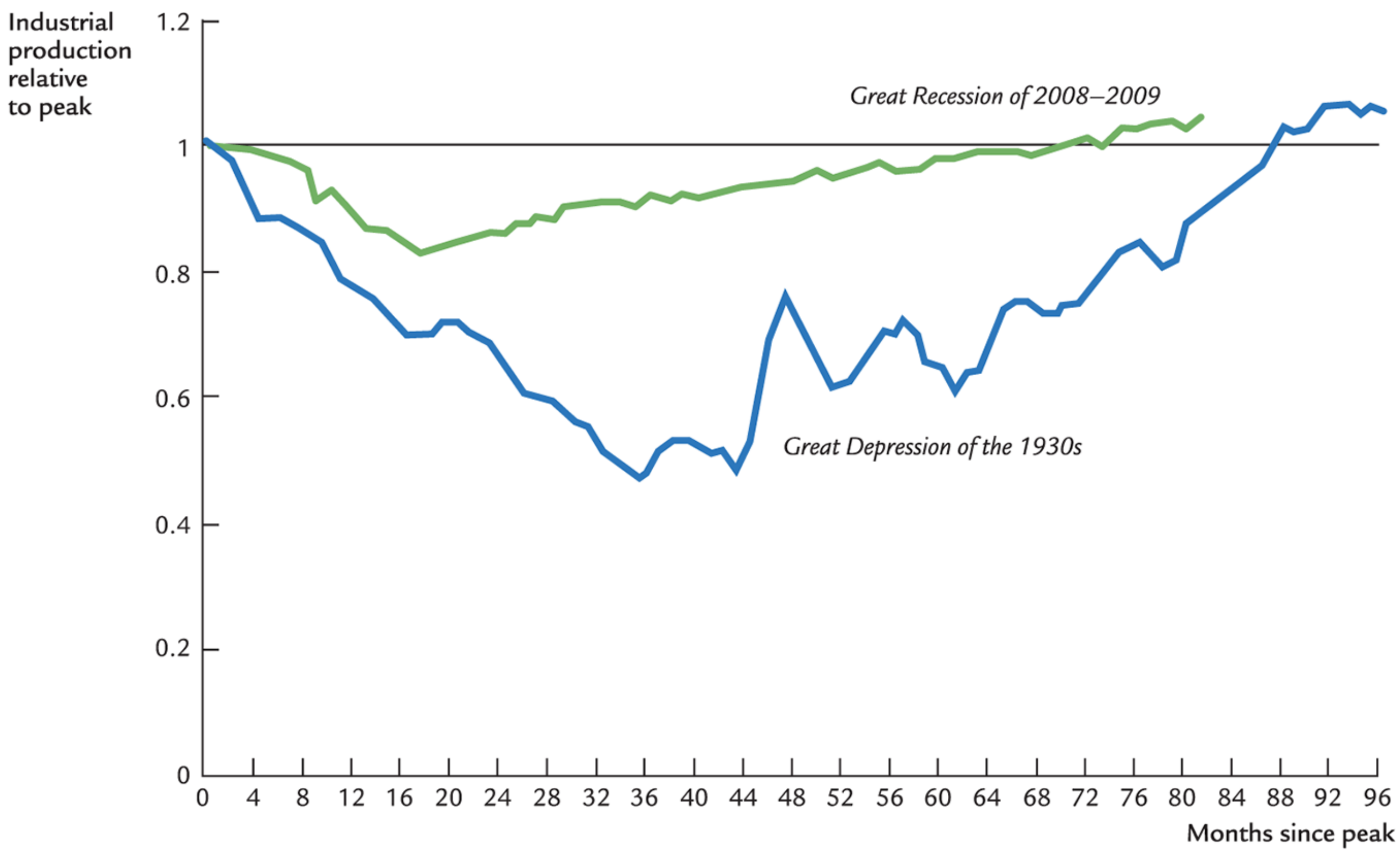


FIGURE 12-9 The Great Recession and the Great Depression This figure compares industrial production during the Great Recession of 2008–2009 and during the Great Depression of the 1930s. Output is normalized to be 100 for the peak before the downturn (December 2007 and August 1929). The data show that the recent downturn was much shallower and shorter.

Data from: Board of Governors of the Federal Reserve System.

This comparison, however, gives only limited comfort. Even though the Great Recession of 2008–2009 was shorter and less severe than the Great Depression, it was nonetheless a devastating event for many families. ■

The Liquidity Trap (Also Known as the Zero Lower Bound)

In the United States in the 1930s, interest rates reached very low levels. As Table 12-1 shows, U.S. interest rates were well under 1 percent throughout the second half of the 1930s. A similar situation occurred during the Great Recession of 2008–2009. In December 2008, the Fed cut its target for the federal funds rate to the range of zero to 0.25 percent, and it kept the rate at that level until 2016.

Some economists describe this situation as a **liquidity trap**. According to the *IS–LM* model, expansionary monetary policy works by reducing interest rates and stimulating investment spending. But if interest rates have already fallen almost to zero, then perhaps monetary policy is no longer effective. Nominal interest rates cannot fall below zero: rather than make a loan at a negative nominal interest rate, a person would just hold cash. In this environment, expansionary monetary policy increases the supply of money, making the public's asset portfolio more liquid, but because interest rates can't fall any further, the extra liquidity might not have any effect. Aggregate demand, production, and employment may be "trapped" at low levels. The liquidity trap is sometimes called the problem of the *zero lower bound*.

Other economists are skeptical about the relevance of liquidity traps and believe that central banks continue to have tools to expand the economy, even after the interest rate target hits the lower bound of zero. One possibility is that a central bank could try to lower longer-term interest rates. It can accomplish this by committing to keep the target interest rate (typically a very short-term interest rate) low for an extended period. A policy of announcing future monetary actions is sometimes called *forward guidance*. A central bank can also lower longer-term interest rates by conducting expansionary open-market operations in a larger variety of financial instruments than it normally does. For example, it could buy long-term government bonds, mortgages, and corporate debt and thereby lower the interest rates on these kinds of loans, a policy sometimes called *quantitative easing*. During the Great Recession and its aftermath, the Fed pursued a policy of both forward guidance and quantitative easing.

Some economists have suggested that the possibility of a liquidity trap argues for a target rate of inflation greater than zero. Under zero inflation, the real interest rate, like the nominal interest, can never fall below zero. But if the normal rate of inflation is, say, 4 percent, then the central bank can push the real interest rate to negative 4 percent by lowering the nominal interest rate toward zero. Put differently, a higher target for the inflation rate means a higher nominal interest rate in normal times (recall the Fisher effect), giving the central bank more room to cut interest rates when the economy experiences recessionary shocks. Thus, a higher inflation target gives monetary policymakers greater scope to stimulate the economy when needed, reducing the likelihood that the economy will hit the zero lower bound and fall into a liquidity trap.[5]

FYI

The Curious Case of Negative Interest Rates

Economists normally think that zero is the lower bound for interest rates. After all, why lend someone money at a negative interest rate when you can simply hold cash? Cash pays an interest rate of zero: a dollar today is still a dollar tomorrow. A zero rate of return is better than a negative one.

Yet in recent years a few central banks around the world have tried stimulating their economies by lowering interest rates below zero. For example, in Switzerland in 2017, the three-month interest rate was negative 0.73 percent. This means that if a person lent out 1,000 Swiss francs, three months later he would be repaid only 998 Swiss francs.

How is this possible? The reason is that storing cash is costly. If you are a typical person, keeping 1,000 francs under your mattress is easy. But suppose you are an executive at a company with 1 billion francs to safeguard. Storing so much money is not simple because there is always the risk of theft or physical decay. (In a scene of the TV show *Narcos*, the drug kingpin Pablo Escobar digs up his hidden hoard of cash only to find that it has decomposed into worthlessness.) As a result, you may be happy to pay a small fee to ensure that your money is returned safely. A negative interest rate represents that fee.

There are, however, limits to how negative interest rates can become. If they are too negative, buying secure vaults to store cash becomes attractive. Thus, while the lower bound on interest rates is not precisely zero, interest rates cannot fall much below zero.

12-4 Conclusion

The purpose of this chapter and the previous one has been to deepen our understanding of aggregate demand. We now have the tools to analyze the effects of monetary and fiscal policy in both the long run and the short run. In the long run, prices are flexible, and we use the classical analysis of Parts Two and Three of this book. In the short run, prices are sticky, and we use the *IS–LM* model to examine how changes in policy influence the economy.

The model in this and the previous chapter provides the basic framework for analyzing the economy in the short run, but it is not the whole story. In Chapter 13 we examine how international interactions affect the theory of aggregate demand. In Chapter 14 we examine the theory behind short-run aggregate supply. Subsequent chapters further refine the theory and examine various issues that arise as the theory is applied to formulate macroeconomic policy. The *IS–LM* model presented in this and the previous chapter provides the starting point for this further analysis.

CHAPTER 13

The Open Economy Revisited: The Mundell–Fleming Model and the Exchange-Rate Regime

> The world is still a closed economy, but its regions and countries are becoming increasingly open. . . . The international economic climate has changed in the direction of financial integration, and this has important implications for economic policy.
>
> —Robert Mundell, 1963

When conducting monetary and fiscal policy, policymakers often need to look beyond their own country's borders. Even if domestic prosperity is their objective, they must consider the rest of the world. The international flows of goods and services and of capital can affect an economy in profound ways.

In this chapter we extend our analysis of aggregate demand to include international trade and finance. The model developed here, called the **Mundell–Fleming model**, has been described as "the dominant policy paradigm for studying open-economy monetary and fiscal policy." In 1999, Robert Mundell won the Nobel Prize for his work in open-economy macroeconomics, including this model.[1]

The Mundell–Fleming model is a close relative of the *IS*–*LM* model. Both models stress the interaction between the goods market and the money market. Both assume that the price level is fixed and then show what causes short-run fluctuations in aggregate income (or, equivalently, shifts in the aggregate demand curve). The key difference is that the *IS*–*LM* model assumes a closed economy, whereas the Mundell–Fleming model assumes an open economy. The Mundell–Fleming model extends the short-run model of national income from Chapters 11 and 12 by including the effects of international trade and finance discussed in Chapter 6.

The Mundell–Fleming model makes one important and extreme assumption: it assumes that the economy being studied is a small open economy with perfect capital mobility. That is, the economy can borrow or lend as much as it wants in world financial markets and, as a result, the economy’s interest rate is determined by the world interest rate. Here is how Mundell himself, in his original 1963 article, explained this assumption:

> In order to present my conclusions in the simplest possible way and to bring the implications for policy into sharpest relief, I assume the extreme degree of mobility that prevails when a country cannot maintain an interest rate different from the general level prevailing abroad. This assumption will overstate the case but it has the merit of posing a stereotype towards which international financial relations seem to be heading. At the same time it might be argued that the assumption is not far from the truth in those financial centers, of which Zurich, Amsterdam, and Brussels may be taken as examples, where the authorities already recognize their lessening ability to dominate money market conditions and insulate them from foreign influences. It should also have a high degree of relevance to a country like Canada whose financial markets are dominated to a great degree by the vast New York market.

As we will see, Mundell’s assumption of a small open economy with perfect capital mobility is useful in developing a tractable and illuminating model.[2]

One lesson from the Mundell–Fleming model is that the behavior of an open economy depends on its exchange-rate system. Indeed, the model was first developed to understand how alternative exchange-rate regimes work and how the choice of exchange-rate regime influences the efficacy of monetary and fiscal policy. We begin by assuming that the economy operates with a floating exchange rate. That is, we assume that the central bank allows the exchange rate to adjust to changing economic conditions. We then examine how the economy operates under a fixed exchange rate. After developing the model, we address an important policy question: what exchange-rate system should a nation adopt?

These issues of open-economy macroeconomics have been very much in the news in recent years. Beginning in 2009, various nations in the European Union, most notably Greece, experienced financial difficulties, leading many observers to wonder whether it was wise for much of the continent to have adopted a common currency—the most extreme form of a fixed exchange rate. When each nation has its own currency, monetary policy and the exchange rate can adjust more easily to the changing needs of each nation. Meanwhile, many American policymakers, including Presidents George W. Bush, Barack Obama, and Donald Trump, have at times complained that China did not allow the value of its currency to float freely against the U.S. dollar. They argued that China kept its currency artificially cheap, making its goods more competitive on world markets. The Mundell–Fleming model offers a good starting point for understanding these policy debates.

13-1 The Mundell–Fleming Model

In this section we construct the Mundell–Fleming model, and in the following sections we use it to analyze various policies. The model is built with components from previous chapters, but these pieces are put together in a new way to address a new set of questions.

The Key Assumption: Small Open Economy with Perfect Capital Mobility

Let's begin with the assumption of a small open economy with perfect capital mobility. As we saw in Chapter 6, this assumption means that the interest rate in this economy r is determined by the world interest rate r^*. Mathematically, we write this assumption as

$$r = r^*.$$

This world interest rate is assumed to be exogenously fixed because the economy is small relative to the world economy, allowing it to borrow or lend as much as it wants in world financial markets without affecting the world interest rate.

Although the idea of perfect capital mobility is expressed with a simple equation, it is important not to lose sight of the sophisticated process this equation represents. Imagine that some event occurred that would normally raise the interest rate (such as a decline in domestic saving). In a small open economy, the domestic interest rate might rise by a little bit for a short time, but as soon as it did, foreigners would see the higher interest rate and start lending to this country (by, for instance, buying this country's bonds). The capital inflow would quickly drive the domestic interest rate back toward r^*. Similarly, if any event started to drive the domestic interest rate downward, capital would flow out of the country to earn a higher return abroad, and this capital outflow would drive the domestic interest rate back up to r^*. Hence, the $r = r^*$ equation represents the assumption that the international flow of capital is rapid enough to keep the domestic interest rate equal to the world interest rate.

The Goods Market and the *IS* Curve

The Mundell–Fleming model describes the market for goods and services much as the *IS*–*LM* model does, but it adds a new term for net exports. In particular, the goods market is represented with the following equation:

$$Y = C(Y - T) + I(r) + G + NX(e).$$

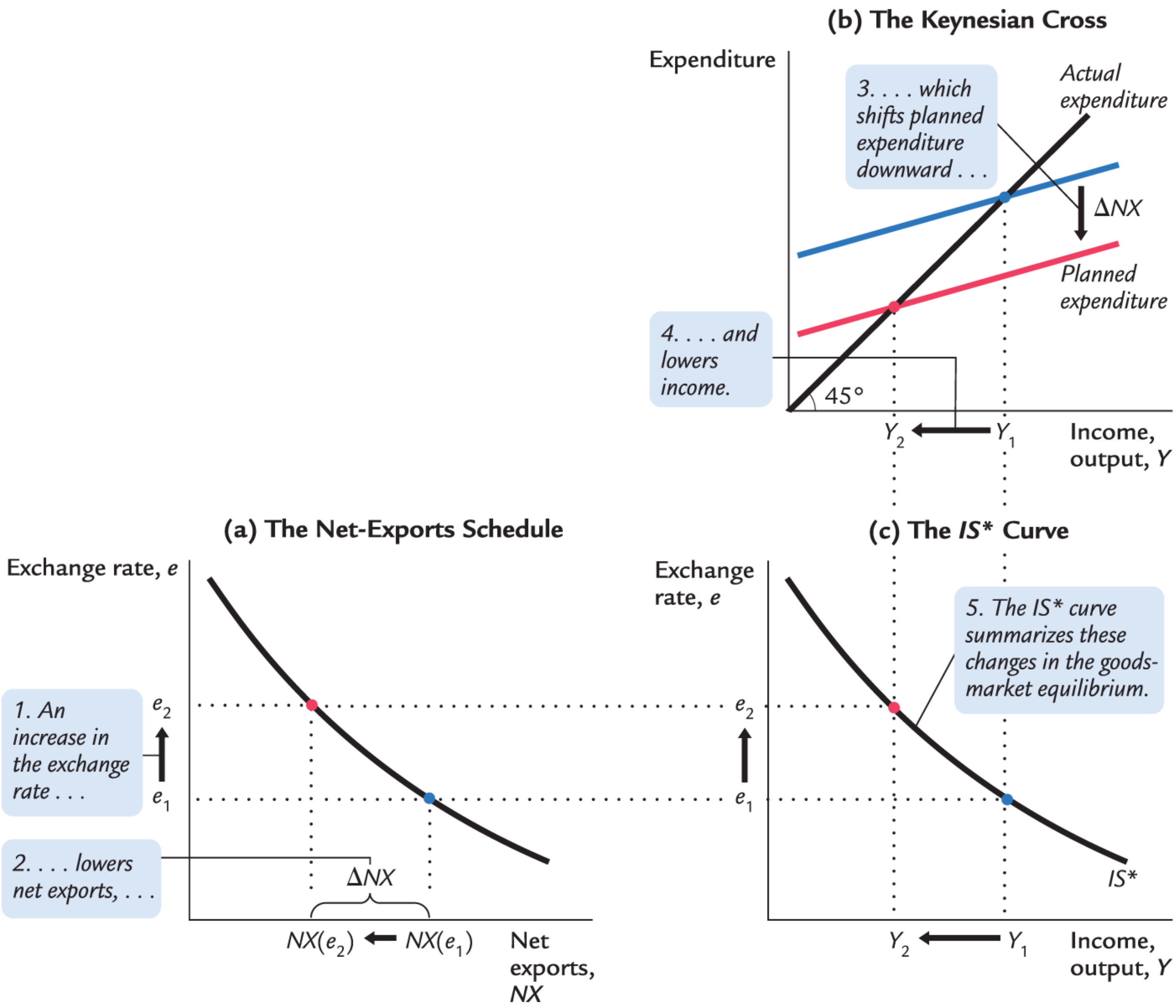
This equation states that aggregate income *Y* is the sum of consumption *C*, investment *I*, government purchases *G*, and net exports *NX*. Consumption depends positively on disposable income $Y - T$. Investment depends negatively on the interest rate. Net exports depend negatively on the exchange rate *e*. As before, we define the exchange rate *e* as the amount of foreign currency per unit of domestic currency; for example, *e* might be 100 yen per dollar.

You may recall that in Chapter 6 we related net exports to the real exchange rate (the relative price of goods at home and abroad) rather than the nominal exchange rate (the relative price of domestic and foreign currencies). If *e* is the nominal exchange rate, then the real exchange rate ε equals eP/P^*, where *P* is the domestic price level and P^* is the foreign price level. The Mundell–Fleming model, however, assumes that the price levels at home and abroad are fixed, so the real exchange rate is proportional to the nominal exchange rate. That is, when the domestic currency appreciates and the nominal exchange rate rises (from, say, 100 to 120 yen per dollar), the real exchange rate rises as well; thus, foreign goods become cheaper compared to domestic goods, causing exports to fall and imports to rise.

The goods market equilibrium condition above has two financial variables that affect expenditure on goods and services (the interest rate and the exchange rate), but we can simplify matters by using the assumption of perfect capital mobility, $r = r^*$:

$$Y = C(Y - T) + I(r^*) + G + NX(e).$$

Let's call this the IS^* equation. The asterisk reminds us that the interest rate is being held constant at the world interest rate r^*. We can illustrate this equation on a graph in which income is on the horizontal axis and the exchange rate is on the vertical axis. This curve is shown in panel (c) of Figure 13-1.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-1 The IS IS^* Curve The IS IS^* curve is derived from the net-exports schedule and the Keynesian cross. Panel (a) shows the net-exports schedule: an increase in the exchange rate from e1 e_1 to e2 e_2 lowers net exports from NX(e1) $NX(e_1)$ to NX(e2). $NX(e_2)$. Panel (b) shows the Keynesian cross: a decrease in net exports from NX(e1) $NX(e_1)$ to NX(e2) $NX(e_2)$ shifts the planned-expenditure schedule downward and reduces income from Y1 Y_1 to Y2. Y_2. Panel (c) shows the IS IS^* curve summarizing this relationship between the exchange rate and income: the higher the exchange rate, the lower the level of income.

The IS IS^* curve slopes downward because a higher exchange rate reduces net exports, which in turn lowers aggregate income. To show how this works, the other panels of Figure 13-1 combine the net-exports schedule and the Keynesian cross to derive the IS IS^* curve. In panel (a), an increase in the exchange rate from e1 e_1 to e2 e_2 lowers net exports from NX(e1) $NX(e_1)$ to NX(e2). $NX(e_2)$. In panel (b), the reduction in net exports shifts the planned-expenditure schedule downward and thus lowers income from Y1 Y_1 to Y2. Y_2. The IS IS^* curve summarizes this relationship between the exchange rate *e* and income *Y*.

The Money Market and the *LM* Curve

The Mundell–Fleming model represents the money market with an equation that should be familiar from the *IS*–*LM* model:

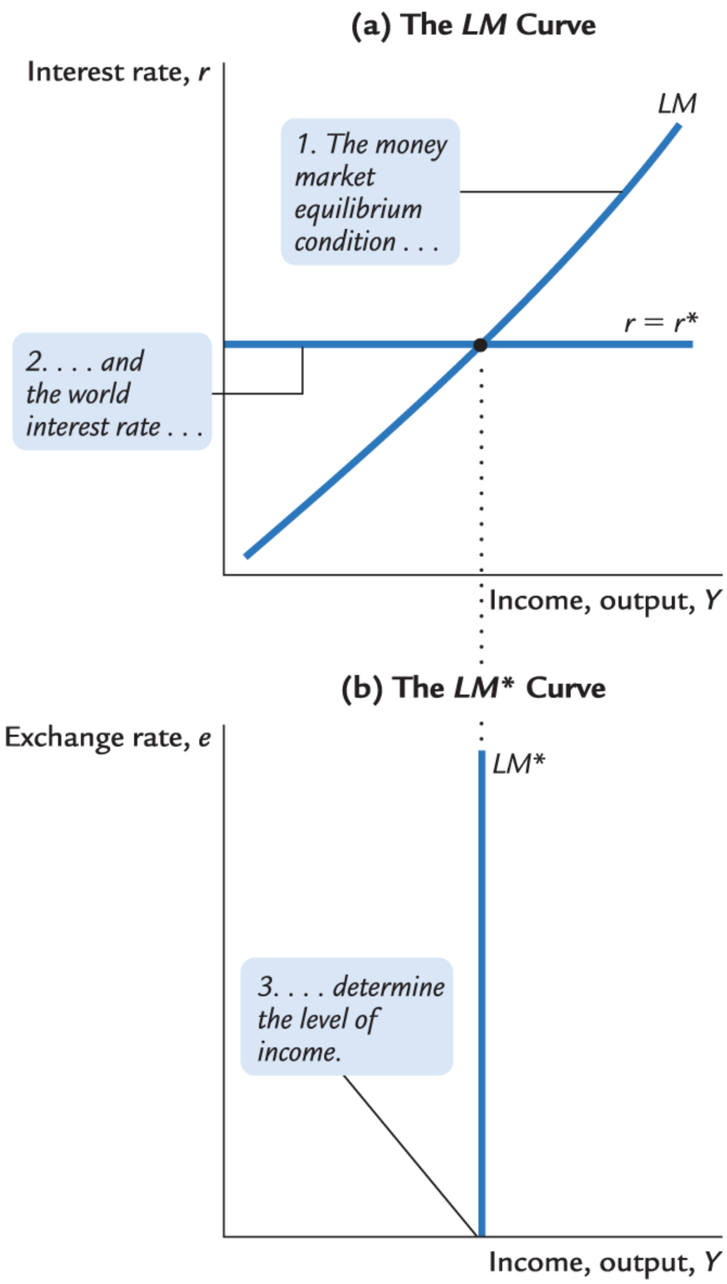
$$M/P = L(r,\ Y).$$

This equation states that the supply of real money balances M/P equals the demand $L(r,\ Y)$. The demand for real balances depends negatively on the interest rate and positively on income. The money supply *M* is an exogenous variable controlled by the central bank, and because the Mundell–Fleming model is designed to analyze short-run fluctuations, the price level *P* is also assumed to be exogenously fixed.

Once again, we add the assumption that the domestic interest rate equals the world interest rate, so $r = r^*$:

$$M/P = L(r^*,\ Y).$$

Let’s call this the LM^* equation. We can represent it graphically with a vertical line, as in panel (b) of Figure 13-2. The LM^* curve is vertical because the exchange rate does not enter the LM^* equation. Given the world interest rate, the LM^* equation determines aggregate income, regardless of the exchange rate. Figure 13-2 shows how the LM^* curve arises from the world interest rate and the *LM* curve, which relates the interest rate and income.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-2 The LM^* Curve Panel (a) shows the standard *LM* curve [which graphs the equation $M/P = L(r, Y)$] together with a horizontal line representing the world interest rate r^*. The intersection of these two curves determines income, regardless of the exchange rate. Therefore, as panel (b) shows, the LM^* curve is vertical.

Putting the Pieces Together

According to the Mundell–Fleming model, a small open economy with perfect capital mobility can be described by two equations:

$$Y = C(Y - T) + I(r^*) + G + NX(e) \quad IS^*,$$
$$M/P = L(r^*, Y) \quad LM^*.$$

$$
\begin{aligned}
Y &= C(Y-T)+I(r^*)+G+NX(e) \qquad IS^*, \\
M/P &= L(r^*,\ Y) \qquad LM^*.
\end{aligned}
$$

The first equation describes equilibrium in the goods market, while the second describes equilibrium in the money market. The exogenous variables are fiscal policy *G* and *T*, monetary policy *M*, the price level *P*, and the world interest rate r r^*. The endogenous variables are income *Y* and the exchange rate *e*.

Figure 13-3 illustrates these two relationships. The equilibrium for the economy is found where the IS IS^* curve and the LM LM^* curve intersect. This intersection shows the exchange rate and income at which the goods market and the money market are both in equilibrium. With this diagram, we can use the Mundell–Fleming model to show how aggregate income *Y* and the exchange rate *e* respond to changes in policy.

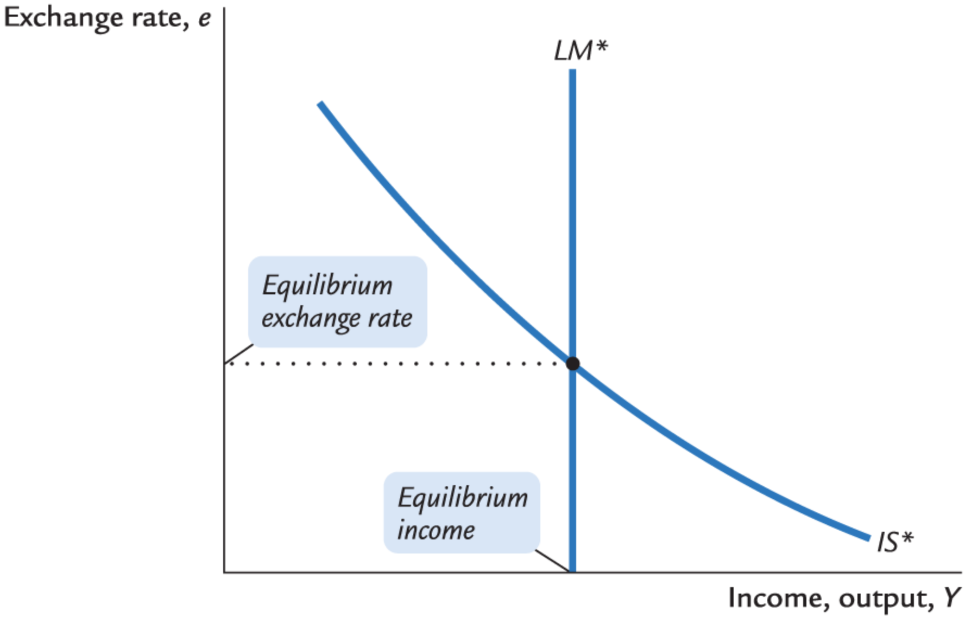


FIGURE 13-3 The Mundell–Fleming Model This graph of the Mundell–Fleming model plots the goods market equilibrium condition IS IS^* and the money market equilibrium condition LM LM^*. Both curves are drawn holding the interest rate constant at the world interest rate. The intersection of these two curves shows the level of income and the exchange rate that satisfy equilibrium both in the goods market and in the money market.

13-2 The Small Open Economy Under Floating Exchange Rates

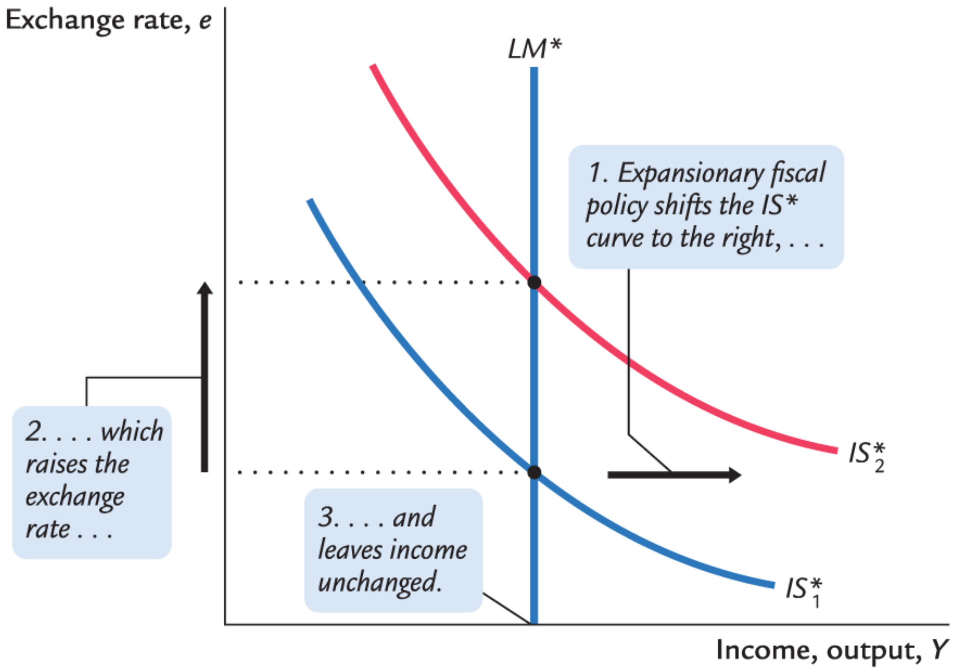
Before analyzing policies in an open economy, we must specify the international monetary system in which the country has chosen to operate. That is, we must consider how people can convert the currency of one country into the currency of another.

We start with the system relevant for most major economies today: **floating exchange rates**. Under a system of floating exchange rates, the exchange rate is set by market forces and can fluctuate in response to changing economic conditions. In this case, the exchange rate *e* adjusts to achieve simultaneous equilibrium in the goods market and the money market. When something happens to change that equilibrium, the exchange rate moves to a new equilibrium value.

Let’s consider three policies that can change the equilibrium: fiscal policy, monetary policy, and trade policy. Our goal is to use the Mundell–Fleming model to show the effects of policy changes and to understand the forces at work as the economy moves from one equilibrium to another.

Fiscal Policy

Suppose the government stimulates domestic spending by increasing government purchases or cutting taxes. Because such expansionary fiscal policy increases planned expenditure, it shifts the IS IS^* curve to the right, as in Figure 13-4. As a result, the exchange rate appreciates, while income remains the same.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-4 A Fiscal Expansion Under Floating Exchange Rates An increase in government purchases or a decrease in taxes shifts the IS IS^* curve to the right. This raises the exchange rate but has no effect on income.

Notice that fiscal policy has very different effects in a small open economy than it does in a closed economy. In the closed-economy *IS–LM* model, a fiscal expansion raises income, whereas in a small open economy with a floating exchange rate, a fiscal expansion leaves income unchanged. Mechanically, the difference arises because the LM LM^* curve is vertical, while the *LM* curve we used to study a closed economy slopes upward. But this explanation is not very satisfying. What are the economic forces that lie behind the different outcomes? To answer this question, we must think through what is happening to the international flow of capital and the implications of these capital flows for the domestic economy.

The interest rate and the exchange rate are the key variables in the story. When income rises in a closed economy, the interest rate rises because higher income increases the demand for money. That is not possible in a small open economy because as soon as the interest rate starts to rise above the world interest rate r r^*, capital flows in from abroad to take advantage of the higher return. As this capital inflow pushes the interest rate back to r r^*, it also has another effect: because foreign investors need to buy the domestic currency to invest in the domestic economy, the capital inflow increases the demand for the domestic currency in the market for foreign-currency exchange, thereby bidding up the value of the domestic currency. The appreciation of the domestic currency makes domestic goods more expensive relative to foreign goods, reducing net exports. The fall in net exports exactly offsets the effects of the expansionary fiscal policy on income.

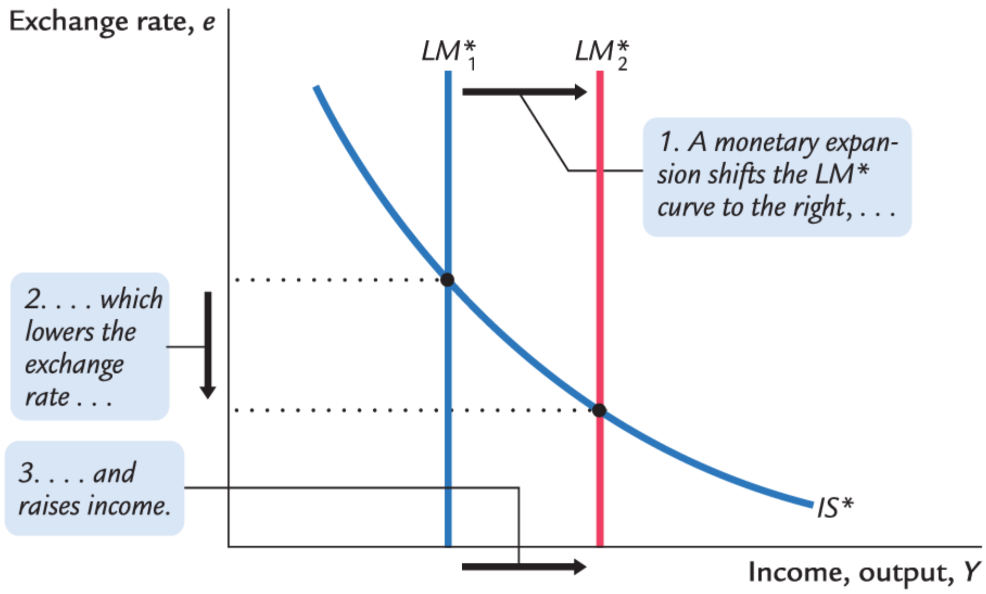
Why is the fall in net exports so great that it renders fiscal policy powerless to influence income? To answer this question, consider the equation that describes the money market:

$$M/P = L(r, Y).$$

In both closed and open economies, the quantity of real money balances supplied M/P is fixed by the central bank (which sets *M*) and the assumption of sticky prices (which fixes *P*). The quantity demanded (determined by *r* and *Y*) must equal this fixed supply. In a closed economy, a fiscal expansion causes the equilibrium interest rate to rise. This increase in the interest rate (which reduces the quantity of money demanded) is accompanied by an increase in equilibrium income (which raises the quantity of money demanded); these two effects together maintain equilibrium in the money market. By contrast, in a small open economy, *r* is fixed at r^*, so there is only one level of income that can satisfy this equation, and this level of income does not change when fiscal policy changes. Thus, when the government increases spending or cuts taxes, the appreciation of the currency and the fall in net exports must be large enough to fully offset the expansionary effect of the policy on income.

Monetary Policy

Suppose now the central bank increases the money supply. Because the price level is assumed to be fixed, the increase in the money supply means an increase in real money balances. The increase in real balances shifts the LM^* curve to the right, as in Figure 13-5. Hence, an increase in the money supply raises income and lowers the exchange rate.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-5 A Monetary Expansion Under Floating Exchange Rates An increase in the money supply shifts the LM^* curve to the right, lowering the exchange rate and raising income.

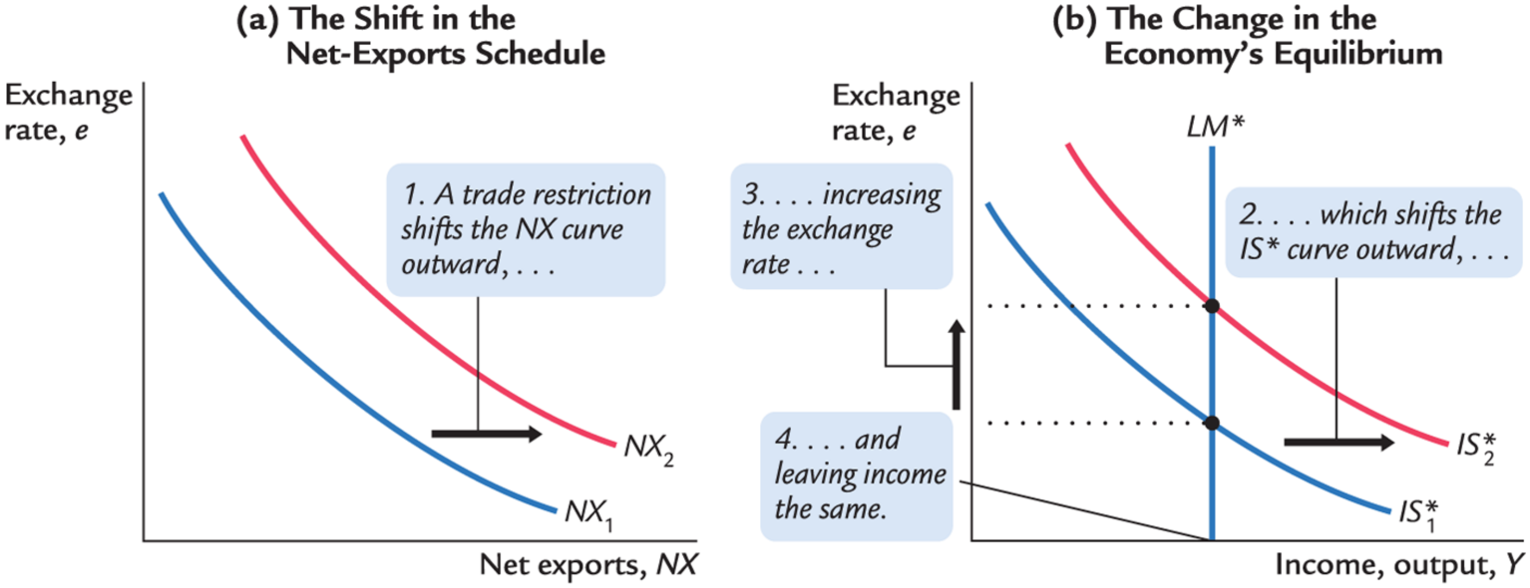
Although monetary policy influences income in an open economy, as it does in a closed economy, the monetary transmission mechanism is different. Recall that in a closed economy, an increase in the money supply increases spending because it lowers the interest rate and stimulates investment. In a small open economy, this channel of monetary transmission is not available because the interest rate is fixed by the world interest rate. So how does monetary policy influence spending? To answer this question, we once again need to think about the international flow of capital and its implications for the domestic economy.

The interest rate and the exchange rate are again the key variables. As soon as an increase in the money supply starts putting downward pressure on the domestic interest rate, capital flows out of the economy because investors seek a higher return elsewhere. This capital outflow prevents the domestic interest rate from falling below the world interest rate r^*. It also has another effect: because investing abroad requires converting domestic currency into foreign currency, the capital outflow increases the supply of the domestic currency in the market for foreign-currency exchange, thereby reducing the value of the domestic currency. This depreciation makes domestic goods less expensive relative to foreign goods, stimulating net exports and thus total income. Hence, in a small open economy, monetary policy influences income by altering the exchange rate rather than the interest rate.

Trade Policy

Suppose the government reduces the demand for imported goods by imposing an import quota or a tariff. What happens to aggregate income and the exchange rate? How does the economy reach its new equilibrium?

Because net exports equal exports minus imports, a reduction in imports means an increase in net exports. That is, the net-exports schedule shifts to the right, as in Figure 13-6. This shift in the net-exports schedule increases planned expenditure and thus moves the IS^* curve to the right. Because the LM^* curve is vertical, the trade restriction raises the exchange rate but does not affect income.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-6 A Trade Restriction Under Floating Exchange Rates A tariff or an import quota shifts the net-exports schedule in panel (a) to the right. As a result, the IS^* curve in panel (b) shifts to the right, raising the exchange rate and leaving income unchanged.

The economic forces behind this transition are similar to the case of expansionary fiscal policy. Because net exports are a component of GDP, the rightward shift in the net-exports schedule, other things equal, puts upward pressure on income *Y*; an increase in *Y*, in turn, increases money demand and puts upward pressure on the interest rate *r*. Foreign capital responds by flowing into the domestic economy, pushing the interest rate back to the world interest rate r^* and increasing the value of the domestic currency. This appreciation makes domestic goods more expensive relative to foreign goods, decreasing net exports *NX* and returning income *Y* to its initial level.

Restrictive trade policies often have the goal of changing the trade balance *NX*. Yet, as we first saw in Chapter 6, such policies do not necessarily have that effect. The same conclusion holds in the Mundell–Fleming model under floating exchange rates. Recall that

$$NX(e) = Y - C(Y - T) - I(r^*) - G.$$

Because a trade restriction does not affect income, consumption, investment, or government purchases, it does not affect the trade balance. Although the shift in the net-exports schedule tends to raise *NX*, the increase in the exchange rate reduces *NX* by the same amount. Thus, the overall effect is simply *less trade*. The domestic economy imports less than it did before the trade restriction, but it exports less as well.

13-3 The Small Open Economy Under Fixed Exchange Rates

We now turn to the second type of exchange-rate system: **fixed exchange rates**. Under a fixed exchange rate, the central bank announces a value for the exchange rate and stands ready to buy and sell the domestic currency to keep the exchange rate at its announced level. This type of system has been used in many historical periods. From 1944 to 1971, most of the world's major economies, including that of the United States, operated within the Bretton Woods system—an international monetary system under which most governments agreed to fix exchange rates. From 1995 to 2005, China fixed the value of its currency against the U.S. dollar—a policy that, as we will see, was a source of some tension between the two countries.

In this section we discuss how such a system works and how policies affect an economy with a fixed exchange rate. Later in the chapter, we examine the pros and cons of fixed exchange rates.

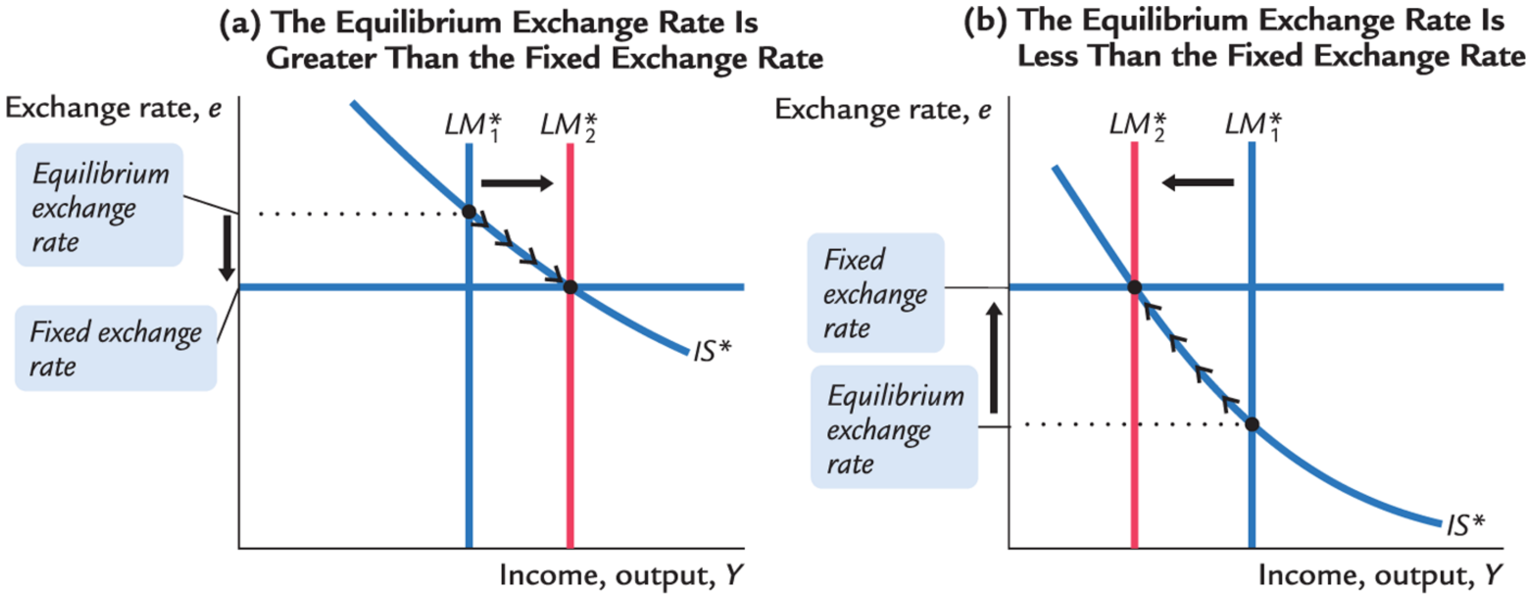
How a Fixed-Exchange-Rate System Works

Under a system of fixed exchange rates, a central bank stands ready to buy or sell the domestic currency for foreign currencies at a predetermined price. For example, suppose the Fed announced that it was going to fix the yen/dollar exchange rate at 100 yen per dollar. It would then stand ready to give $1 in exchange for 100 yen or to give 100 yen in exchange for $1. To carry out this policy, the Fed would need a reserve of dollars (which it can print) and a reserve of yen (which it must have purchased previously).

A fixed exchange rate dedicates a country's monetary policy to the single goal of keeping the exchange rate at the announced level. In other words, the essence of a fixed-exchange-rate system is the commitment of the central bank to allow the money supply to adjust to whatever level will ensure that the equilibrium exchange rate in the market for foreign-currency exchange equals the announced exchange rate. Moreover, as long as the central bank stands ready to buy or sell foreign currency at the fixed exchange rate, the money supply adjusts automatically to the necessary level.

To see how fixing the exchange rate determines the money supply, consider an example. Suppose the Fed decides to fix the exchange rate at 100 yen per dollar, but, in the current equilibrium with the current money supply, the market exchange rate is 150 yen per dollar. This situation is shown in panel (a) of Figure 13-7. Notice that there is a profit opportunity: an arbitrageur could buy 300 yen in the foreign-exchange market for $2 and then sell the yen to the Fed for $3, making a $1 profit. When the Fed buys these yen from the arbitrageur, the dollars it pays for them increase the money supply. The rise in the money supply shifts the

LM LM^* curve to the right, lowering the equilibrium exchange rate. In this way, the money supply continues to rise until the equilibrium exchange rate falls to the level the Fed has announced.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-7 How a Fixed Exchange Rate Governs the Money Supply In panel (a), the equilibrium exchange rate initially exceeds the fixed level. Arbitrageurs will buy foreign currency in foreign-exchange markets and sell it to the Fed for a profit. This process increases the money supply, shifting the LM LM^* curve to the right and lowering the exchange rate. In panel (b), the equilibrium exchange rate is initially below the fixed level. Arbitrageurs will buy foreign currency from the Fed and sell it in foreign-exchange markets for a profit. This process reduces the money supply, shifting the LM LM^* curve to the left and raising the exchange rate.

Conversely, suppose that when the Fed decides to fix the exchange rate at 100 yen per dollar, the equilibrium has a market exchange rate of 50 yen per dollar. Panel (b) of Figure 13-7 shows this situation. In this case, an arbitrageur could make a profit by buying 100 yen from the Fed for $1 and then selling the yen in the marketplace for $2. When the Fed sells these yen, the $1 it receives reduces the money supply. The fall in the money supply shifts the LM LM^* curve to the left, raising the equilibrium exchange rate. The money supply continues to fall until the equilibrium exchange rate rises to the announced level.

Note that this exchange-rate system fixes the *nominal* exchange rate. Whether it also fixes the real exchange rate depends on the time horizon under consideration. If prices are flexible, as they are in the long run, the real exchange rate can change even while the nominal exchange rate is fixed. Therefore, in the long run described in Chapter 6, a policy to fix the nominal exchange rate would not influence any real variable, including the real exchange rate. A fixed nominal exchange rate would influence only the money supply and the price level. Yet in the short run described by the Mundell–Fleming model, prices are fixed, so a fixed nominal exchange rate implies a fixed real exchange rate as well.

CASE STUDY

The International Gold Standard

During the late nineteenth and early twentieth centuries, most of the world's major economies operated under the gold standard. Each country maintained a reserve of gold and agreed to exchange one unit of its currency for a specified amount of gold. Through the gold standard, the world’s economies maintained a system of fixed exchange rates.

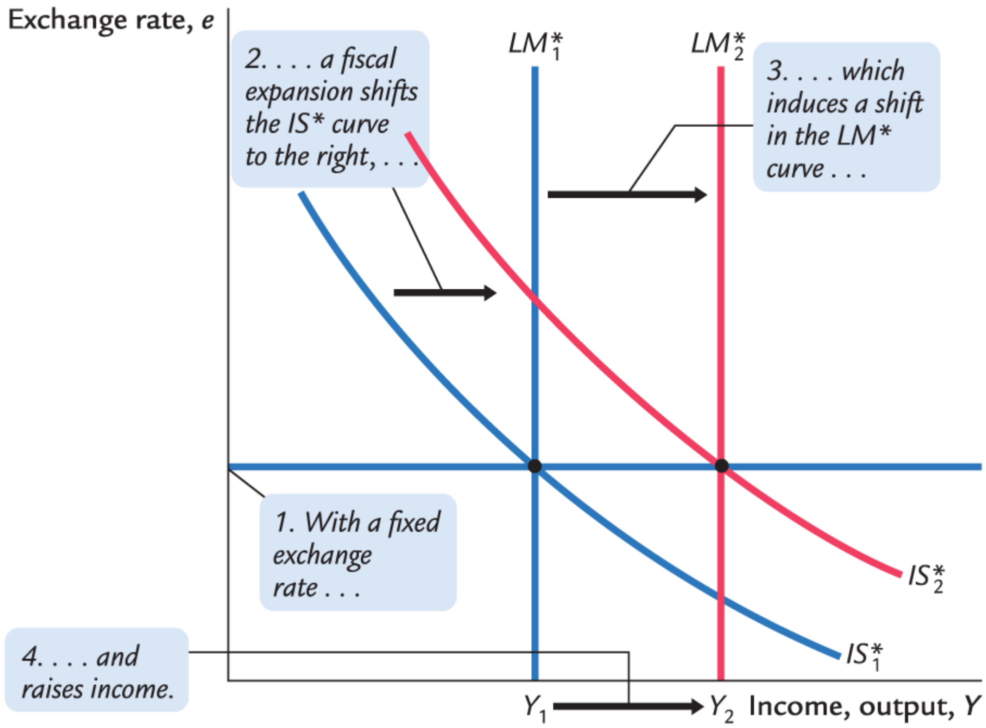
To see how an international gold standard fixes exchange rates, suppose the U.S. Treasury stands ready to buy or sell 1 ounce of gold for $100, and the Bank of England stands ready to buy or sell 1 ounce of gold for 100 pounds. Together, these policies fix the rate of exchange between dollars and pounds: $1 must trade for 1 pound. Otherwise, the law of one price would be violated, and it would be profitable to buy gold in one country and sell it in the other.

For example, suppose the market exchange rate is 2 pounds per dollar. In this case, an arbitrageur could buy 200 pounds for $100, use the pounds to buy 2 ounces of gold from the Bank of England, bring the gold to the United States, and sell it to the Treasury for $200—making a $100 profit. Moreover, by bringing the gold to the United States from England, the arbitrageur would increase the money supply in the United States and decrease the money supply in England.

Thus, during the era of the gold standard, the international transport of gold by arbitrageurs was an automatic mechanism adjusting the money supply and stabilizing exchange rates. This system did not completely fix exchange rates because shipping gold across the Atlantic was costly. Yet the international gold standard did keep the exchange rate within a range dictated by transportation costs. It thereby prevented large and persistent movements in exchange rates. [3] ■

Fiscal Policy

Let’s now examine how economic policies affect a small open economy with a fixed exchange rate. Suppose the government stimulates domestic spending by increasing government purchases or cutting taxes. This policy shifts the *IS* curve to the right, as in Figure 13-8, putting upward pressure on the market exchange rate. But because the central bank stands ready to trade foreign and domestic currency at the fixed exchange rate, arbitrageurs respond to the rising exchange rate by selling foreign currency to the central bank, causing an automatic monetary expansion. The rise in the money supply shifts the LM LM^* curve to the right. Thus, under a fixed exchange rate, a fiscal expansion raises aggregate income.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-8 A Fiscal Expansion Under Fixed Exchange Rates A fiscal expansion shifts the IS IS^* curve to the right. To maintain the fixed exchange rate, the Fed increases the money supply, thereby shifting the LM LM^* curve to the right. Hence, in contrast to the case of floating exchange rates, under fixed exchange rates a fiscal expansion raises income.

Monetary Policy

Imagine that a central bank operating with a fixed exchange rate tries to increase the money supply—for example, by buying bonds from the public. What would happen? The initial impact of this policy is to shift the LM LM^* curve to the right, lowering the exchange rate, as in Figure 13-9. But, because the central bank is committed to trading foreign and domestic currency at a fixed exchange rate, arbitrageurs respond to the falling exchange rate by selling the domestic currency to the central bank, causing the money supply and the LM LM^* curve to return to their initial positions. Hence, monetary policy as usually conducted is ineffectual under a fixed exchange rate. By agreeing to fix the exchange rate, the central bank gives up its control over the money supply.

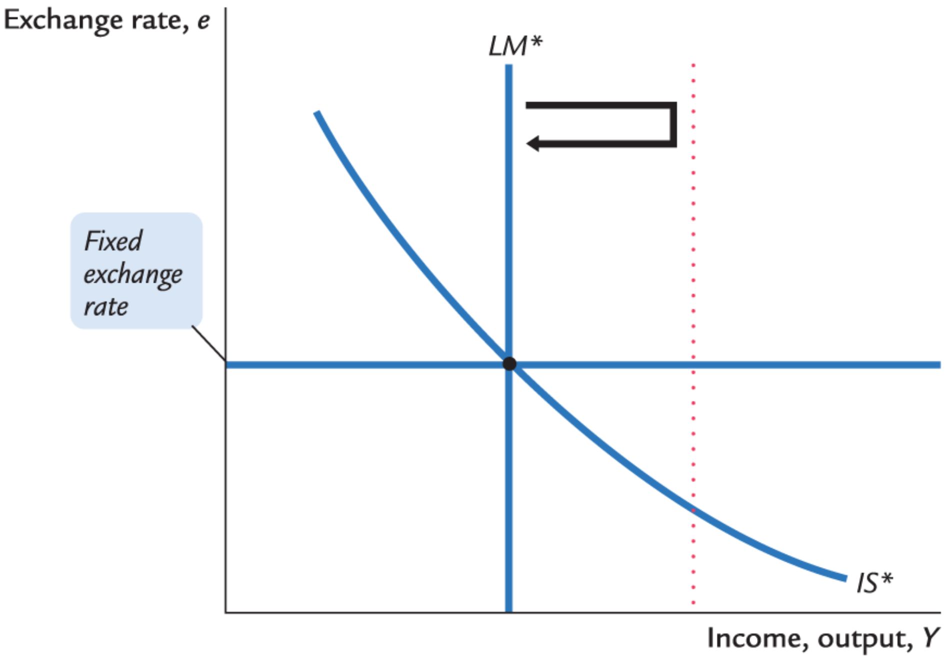


FIGURE 13-9 A Monetary Expansion Under Fixed Exchange Rates If the Fed tries to increase the money supply—for example, by buying bonds from the public—it will put downward pressure on the exchange rate. To maintain the fixed exchange rate, the money supply and the LM LM^* curve must return to their initial positions. Hence, under fixed exchange rates, normal monetary policy is ineffectual.

A country with a fixed exchange rate can, however, conduct a type of monetary policy: it can decide to change the level at which the exchange rate is fixed. A reduction in the official value of the currency is called a **devaluation**, and an increase in its official value is called a **revaluation**. In the Mundell–Fleming model, a devaluation shifts the LM LM^* curve to the right; it acts like an increase in the money supply under a floating exchange rate. A devaluation thus expands net exports and raises aggregate income. Conversely, a revaluation shifts the LM LM^* curve to the left, reduces net exports, and lowers aggregate income.

CASE STUDY

Devaluation and the Recovery from the Great Depression

The Great Depression of the 1930s was a global problem. Although events in the United States may have precipitated the downturn, all of the world's major economies experienced huge declines in production and employment. Yet not all governments responded to this calamity in the same way.

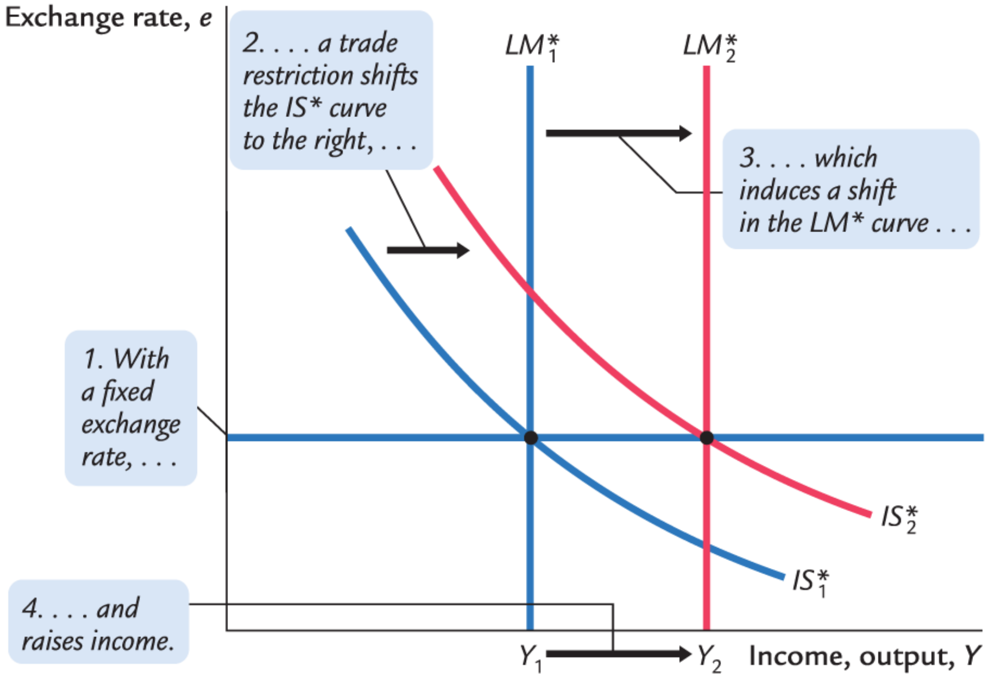
One key difference among governments was how committed they were to the fixed exchange rate set by the international gold standard. Some countries, such as France, Germany, Italy, and the Netherlands, maintained the old rate of exchange between gold and currency. Other countries, such as Denmark, Finland, Norway, Sweden, and the United Kingdom, reduced the amount of gold they would pay for each unit of currency by about 50 percent. By reducing the gold content of their currencies, these governments devalued their currencies relative to those of other countries.

The subsequent experience of these two groups of countries confirms the prediction of the Mundell–Fleming model. Those countries that pursued a policy of devaluation recovered quickly from the Depression. The lower value of the currency increased the money supply, stimulated exports, and expanded production. By contrast, those countries that maintained the old exchange rate suffered longer with a depressed level of economic activity.

What about the United States? President Herbert Hoover kept the United States on the gold standard, but in a controversial move, President Franklin Roosevelt took the nation off it in June 1933, just three months after taking office. That date roughly coincides with the end of the deflation and the beginning of recovery. Many economic historians believe that removing the nation from the gold standard was the most significant policy action that President Roosevelt took to end the Great Depression.[4] ■

Trade Policy

Suppose the government reduces imports by imposing an import quota or a tariff. This policy shifts the net-exports schedule to the right and thus shifts the IS^* curve to the right, as in Figure 13-10. The shift in the IS^* curve tends to raise the exchange rate. To keep the exchange rate at the fixed level, the money supply must rise, shifting the LM^* curve to the right.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-10 A Trade Restriction Under Fixed Exchange Rates A tariff or an import quota shifts the IS^* curve to the right. This induces an increase in the money supply to maintain the fixed exchange rate. Hence, aggregate income increases.

The result of a trade restriction under a fixed exchange rate is very different from that under a floating exchange rate. In both cases, a trade restriction shifts the net-exports schedule to the right, but only under a fixed exchange rate does a trade restriction increase net exports NX. The reason is that a trade restriction under a fixed exchange rate induces monetary expansion rather than an appreciation of the currency. The monetary expansion, in turn, raises aggregate income. Recall the accounting identity

$$NX = S - I.$$

When income rises, saving also rises, and this implies an increase in net exports.

Policy in the Mundell–Fleming Model: A Summary

The Mundell–Fleming model shows that the effect of almost any policy on a small open economy depends on whether the exchange rate is floating or fixed. Table 13-1 summarizes our analysis of the short-run effects of fiscal, monetary, and trade policies on income, the exchange rate, and the trade balance. Note that all of the results are different under floating and fixed exchange rates.

TABLE 13-1 The Mundell–Fleming Model: Summary of Policy Effects

	EXCHANGE-RATE REGIME					
	FLOATING			FIXED		
	IMPACT ON:					
Policy	***Y***	***e***	***NX***	***Y***	***e***	***NX***
Fiscal expansion	0	↑	↓	↑	0	0
Monetary expansion	↑	↓	↑	0	0	0
Import restriction	0	↑	0	↑	0	↑

Note: This table shows the direction of impact of various economic policies on income *Y*, the exchange rate *e*, and the trade balance *NX*. A "↑" indicates that the variable increases; a "↓" indicates that it decreases; a "0" indicates no effect. Remember that the exchange rate is defined as the amount of foreign currency per unit of domestic currency (for example, 100 yen per dollar).

In particular, the Mundell–Fleming model shows that the power of monetary and fiscal policy to influence aggregate income depends on the exchange-rate regime. Under floating exchange rates, only monetary policy affects income. The usual expansionary impact of fiscal policy is offset by an appreciation of the currency and a decrease in net exports. Under fixed exchange rates, only fiscal policy affects income. The normal potency of monetary policy is lost because the money supply is dedicated to maintaining the exchange rate at the announced level.

13-4 Interest Rate Differentials

So far, our analysis has assumed that the interest rate in a small open economy equals the world interest rate: r=r $r = r^*$. To some extent, however, interest rates differ around the world. We now extend our analysis by considering the causes and effects of international interest rate differentials.

Country Risk and Exchange-Rate Expectations

When we assumed earlier that the interest rate in our small open economy is determined by the world interest rate, we were applying the law of one price. We reasoned that if the domestic interest rate was above the world interest rate, people from abroad would lend to that country, driving down the domestic interest rate. And if the domestic interest rate was below the world interest rate, domestic residents would lend abroad to earn a higher return, driving up the domestic interest rate. In the end, the domestic interest rate equals the world interest rate.

Why doesn't this logic always apply? There are two reasons.

One reason is country risk. When investors buy U.S. government bonds or make loans to U.S. corporations, they are fairly confident that they will be repaid with interest. By contrast, in some less-developed countries, it is plausible to fear that a revolution or some other political upheaval might lead to a default on loan repayments. Borrowers in such countries often have to pay higher interest rates to compensate lenders for this risk.

Another reason interest rates differ across countries is expected changes in the exchange rate. For example, suppose people expect the Mexican peso to depreciate relative to the U.S. dollar. Then loans made in pesos will be repaid in a less valuable currency than loans made in dollars. To compensate for this expected fall in the Mexican currency, the interest rate in Mexico will be higher than the interest rate in the United States.

Thus, because of country risk and expectations about future exchange-rate changes, the interest rate of a small open economy can differ from interest rates in other economies around the world. Let's see how this fact affects our analysis.

Differentials in the Mundell–Fleming Model

Consider again the Mundell–Fleming model with a floating exchange rate. To incorporate interest rate differentials into the model, we assume that the interest rate in our small open economy is determined by the world interest rate plus a risk premium θ:

$$r = r^* + \theta.$$

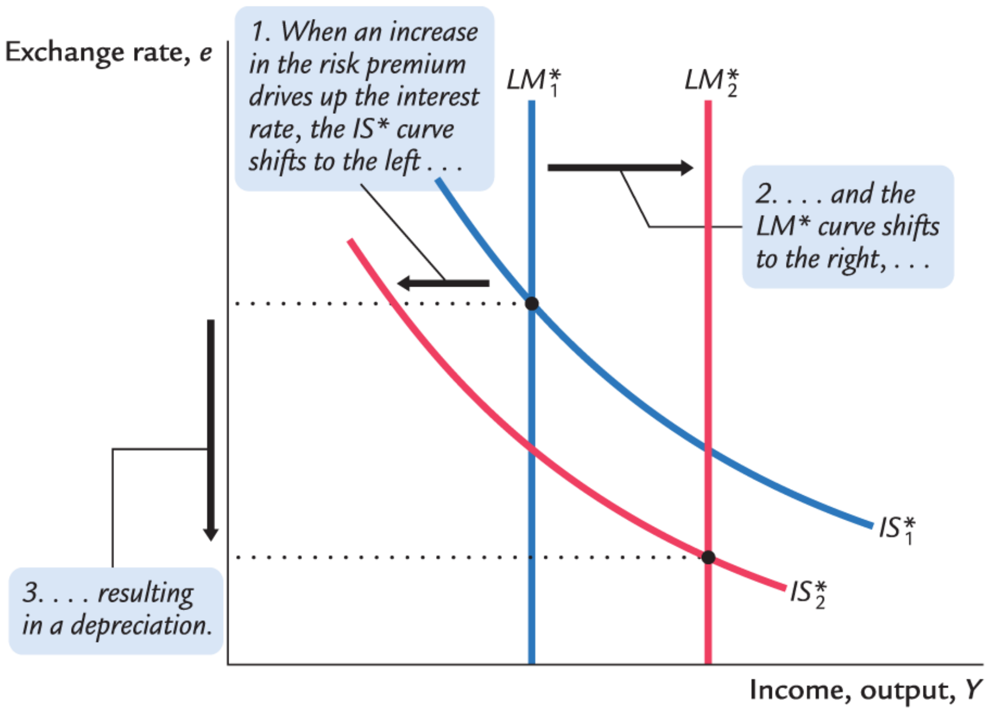
The risk premium is determined by the perceived political risk of making loans in a country and the expected change in the real exchange rate. For our purposes here, we can take the risk premium as exogenous in order to examine how changes in the risk premium affect the economy.

The model is largely the same as before. The two equations are

$$\begin{aligned} Y &= C(Y - T) + I(r^* + \theta) + G + NX(e) \qquad IS^*, \\ M/P &= L(r^* + \theta,\ Y) \qquad LM^*. \end{aligned}$$

For any given fiscal policy, monetary policy, price level, and risk premium, these two equations determine the level of income and exchange rate that equilibrate the goods market and the money market. Holding constant the risk premium, the tools of monetary, fiscal, and trade policy work just as we have already seen.

Now suppose political turmoil causes the country’s risk premium θ to rise. Because $r = r^* + \theta$, the most direct effect is that the domestic interest rate *r* rises. The higher interest rate, in turn, has two effects. First, the IS^* curve shifts to the left because the higher interest rate reduces investment. Second, the LM^* curve shifts to the right because the higher interest rate reduces the demand for money, implying higher income for any given money supply. [Recall that *Y* must satisfy the equation $M/P = L(r^* + \theta,\ Y)$.] As Figure 13-11 shows, these shifts cause income to rise and the currency to depreciate.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-11 An Increase in the Risk Premium An increase in the risk premium associated with a country drives up its interest rate. Because the higher interest rate reduces investment, the IS IS^* curve shifts to the left. Because it also reduces money demand, the LM LM^* curve shifts to the right. Income rises, and the currency depreciates.

This analysis has an important implication: expectations about the exchange rate are partially self-fulfilling. For example, suppose that for some reason people come to believe that the Mexican peso will be worth less in the future. Because of this belief, investors will place a larger risk premium on Mexican assets (θ θ will increase in Mexico). Mexican interest rates will rise and, as we have just seen, the value of the Mexican currency will fall. *Thus, the expectation that a currency will lose value in the future causes it to lose value today.*

One surprising—and perhaps inaccurate—prediction of this analysis is that an increase in country risk as measured by θθ will cause the economy's income to increase. This occurs in Figure 13-11 because of the rightward shift in the LM LM^* curve. Although higher interest rates depress investment, the depreciation of the currency stimulates net exports by an even greater amount. As a result, aggregate income rises.

In practice, however, such a boom in income typically does not occur—for three reasons. First, the central bank might want to avoid the large depreciation of the domestic currency and, therefore, may respond by decreasing the money supply *M*. Second, the depreciation of the domestic currency may increase the price of imported goods, which raises the price level *P*. Third, when some event increases the country risk premium θ, θ, residents of the country might respond to the event by increasing their demand for money (for any given income and interest rate) because money is often the safest asset available. All three of these changes shift the LM LM^* curve toward the left, which mitigates the fall in the exchange rate but also depresses income.

Thus, increases in country risk are not desirable. In the short run, they lead to a depreciating currency and, through the three channels just described, falling aggregate income. In addition, because a higher interest rate reduces investment, the long-run implications are reduced capital accumulation and lower economic growth.

CASE STUDY

International Financial Crisis: Mexico 1994–1995

In August 1994, a Mexican peso was worth 30 cents. A year later, it was worth only 16 cents. What explains this massive fall in the value of the Mexican currency? Country risk is a large part of the story.

At the beginning of 1994, Mexico was a country on the rise. The recent passage of the North American Free Trade Agreement (NAFTA), which reduced trade barriers among the United States, Canada, and Mexico, made many people confident about the future of the Mexican economy. Investors around the world were eager to make loans to the Mexican government and to Mexican corporations.

Political developments soon changed that perception. A violent uprising in the Chiapas region of Mexico made the political situation in Mexico seem precarious. Then Luis Donaldo Colosio, the leading presidential candidate, was assassinated. The political future looked less certain, and many investors started placing a larger risk premium on Mexican assets.

At first, the rising risk premium did not affect the value of the peso because Mexico was operating with a fixed exchange rate. As we have seen, under a fixed exchange rate, the central bank agrees to trade the domestic currency (pesos) for a foreign currency (dollars) at a predetermined rate. Thus, when an increase in the country risk premium put downward pressure on the value of the peso, the Mexican central bank had to accept pesos and pay out dollars. This exchange-market intervention contracted the Mexican money supply (shifting the LM LM^* curve to the left) when the currency might otherwise have depreciated.

Yet Mexico’s foreign-currency reserves were too small to maintain its fixed exchange rate. When Mexico ran out of dollars at the end of 1994, the Mexican government devalued the peso. This decision had repercussions, however, because the government had promised that it would not devalue. Investors became even more distrustful of Mexican policymakers and feared further devaluations.

Investors around the world (including those in Mexico) avoided buying Mexican assets. The country risk premium rose once again, adding to the upward pressure on interest rates and the downward pressure on the peso. The Mexican stock market plummeted. When the Mexican government needed to roll over some of its debt that was coming due, investors were unwilling to buy the new debt. Default appeared to be the government’s only option. In just a few months, Mexico had gone from being a promising emerging economy to being a risky economy with a government on the verge of bankruptcy.

Then the United States stepped in. The U.S. government had three motives: to help its neighbor to the south, to prevent the massive illegal immigration that might follow government default and economic collapse, and to prevent the investor pessimism regarding Mexico from spreading to other developing countries. The U.S. government, together with the International Monetary Fund (IMF), led an international effort to bail out the Mexican government. In particular, the United States provided loan guarantees for Mexican government debt, which allowed the Mexican government to refinance the debt that was coming due. These loan guarantees helped restore confidence in the Mexican economy, thereby reducing to some extent the country risk premium.

Although the U.S. loan guarantees may well have stopped a bad situation from getting worse, they did not prevent the Mexican meltdown of 1994–1995 from being a painful experience for the Mexican people. The peso lost much of its value, and Mexico went through a deep recession. Fortunately, by the late 1990s, the worst was over, and aggregate income was growing again.

The lesson from this experience is clear: changes in perceived country risk, often attributable to political instability, are an important determinant of interest rates and exchange rates in small open economies. ■

CASE STUDY

International Financial Crisis: Asia 1997–1998

In 1997, as the Mexican economy was recovering from its financial crisis, a similar story started to unfold in several Asian economies, including those of Thailand, South Korea, and especially Indonesia. The symptoms were familiar: high interest rates, falling asset values, and a depreciating currency. In Indonesia, for instance, short-term nominal interest rates rose above 50 percent, the stock market lost about 90 percent of its value (measured in U.S. dollars), and the rupiah fell against the dollar by more than 80 percent. The crisis led to rising inflation in these countries (because the depreciating currency made imports more expensive) and to falling GDP (because high interest rates and reduced confidence depressed spending). Real GDP in Indonesia fell about 13 percent in 1998.

What sparked this firestorm? The problem began in the Asian banking systems. For many years, the governments in the Asian countries had been more involved in managing the allocation of resources—in particular, financial resources—than is true in the United States and other developed countries. Some commentators had applauded this “partnership” between government and private enterprise, even suggesting that the United States should follow the example. Over time, however, it became clear that many Asian banks had been extending loans to those with the most political clout rather than to those with the most profitable investment projects. Once rising default rates started to expose this “crony capitalism,” as it was then called, international investors started to lose confidence in the future of these economies. The risk premiums for Asian assets rose, causing interest rates to skyrocket and currencies to collapse.

International crises of confidence often involve a vicious circle that can amplify the problem. Here is a brief account about what happened in Asia:

1. Problems in the banking system eroded international confidence in these economies.
2. Loss of confidence raised risk premiums and interest rates.
3. Rising interest rates, together with the loss of confidence, depressed the prices of stock and other assets.
4. Falling asset prices reduced the value of collateral being used for bank loans.
5. Reduced collateral increased default rates on bank loans.
6. Greater defaults exacerbated problems in the banking system. Now return to step 1 to complete and continue the circle.

Some economists have used this vicious-circle argument to suggest that the Asian crisis was a self-fulfilling prophecy: bad things happened because people expected bad things to happen. Most economists, however, thought the political corruption of the banking system was a real problem, which was then compounded by this vicious circle of reduced confidence.

Exacerbating the situation was a *currency mismatch* between the assets and liabilities of financial institutions.

Banks in these emerging economies often borrowed from abroad in foreign currencies, such as the U.S. dollar, and made loans to residents of their own countries in their domestic currencies, such as the rupiah. As a result, they had assets denominated in the domestic currency but liabilities denominated in a foreign currency. When the domestic currency depreciated in foreign-exchange markets, the value of the banks' assets fell relative to their liabilities, making the problems of the banking system even worse.

As the Asian crisis developed, the IMF and the United States tried to restore confidence, much as they had with Mexico a few years earlier. In particular, the IMF made loans to the Asian countries to help them through the crisis; in exchange for these loans, it exacted promises that the governments would reform their banking systems and eliminate crony capitalism. The IMF's hope was that the short-term loans and longer-term reforms would restore confidence, lower the risk premium, and turn the vicious circle into a virtuous one. This policy seems to have worked: the Asian economies recovered quickly from their crisis. ■

13-5 Should Exchange Rates Be Floating or Fixed?

Having seen how an economy works under floating and fixed exchange rates, let's consider which exchange-rate regime is better.

Pros and Cons of Different Exchange-Rate Systems

"Then it's agreed. Until the dollar firms up, we let the clamshell float."

The main argument for a floating exchange rate is that it allows a nation to use its monetary policy for other purposes. Under fixed rates, monetary policy is committed to the single goal of maintaining the exchange rate at its announced level. Yet the exchange rate is only one of many economic variables that monetary policy can influence. A system of floating exchange rates lets monetary policymakers pursue other goals, such as stabilizing employment or prices.

Advocates of fixed exchange rates argue that exchange-rate uncertainty makes international trade more difficult. After the world abandoned the Bretton Woods system of fixed exchange rates in the early 1970s, both real and nominal exchange rates became (and have remained) much more volatile than anyone had expected. Some economists attribute this volatility to irrational and destabilizing speculation by international investors. Business executives often claim that this volatility is harmful because it increases the uncertainty that accompanies international business transactions. Despite this exchange-rate volatility, however, the amount of world trade has continued to rise under floating exchange rates.

Advocates of fixed exchange rates sometimes argue that a commitment to a fixed exchange rate is one way to discipline a nation's monetary authority and prevent excessive growth in the money supply. Yet there are many other policy rules to which the central bank could be committed. In Chapter 16, for instance, we discuss policy rules such as targets for nominal GDP or the inflation rate. Fixing the exchange rate has the advantage of being simpler to implement than these other policy rules because the money supply adjusts automatically. But this policy may lead to greater volatility in income and employment.

In practice, the choice between floating and fixed rates is not as stark as it may seem at first. Under systems of fixed exchange rates, countries can change the value of their currency if maintaining the exchange rate conflicts too severely with other goals. Under systems of floating exchange rates, countries often use formal or informal targets for the exchange rate when setting monetary policy. We rarely observe exchange rates that are completely fixed or completely floating. Instead, under both systems, stability of the exchange rate is usually one among many objectives of the central bank.

CASE STUDY

The Debate over the Euro

If you have ever driven the 3,000 miles from New York City to San Francisco, you may recall that you never needed to change your money from one form of currency to another. In all 50 U.S. states, local residents are happy to accept the U.S. dollar for the items you buy. Such a *monetary union* is the most extreme form of a fixed exchange rate. The exchange rate between New York dollars and San Francisco dollars is so irrevocably fixed that you may not even know that there is a difference between the two. (What's the difference? Each dollar bill is issued by one of the dozen local Federal Reserve Banks. Although the bank of origin can be identified from the bill's markings, you don't care which type of dollar you hold because everyone else, including the Federal Reserve system, is ready to trade any dollar from one bank for a dollar from another.)

If you made a similar 3,000-mile trip across Europe during the 1990s, however, your experience was very different. You didn't have to travel far before needing to exchange your French francs for German marks, Dutch guilders, Spanish pesetas, or Italian lira. The large number of currencies in Europe made traveling less convenient and more expensive. Every time you crossed a border, you had to wait in line at a bank to get the local money, and you had to pay the bank a fee for the service.

Today, however, the situation in Europe is more like that in the United States. Many European countries have given up having their own currencies and have formed a monetary union that uses a common currency called the *euro*. As a result, the exchange rate between France and Germany is now as fixed as the exchange rate between New York and California.

The introduction of a common currency has its costs. The most important is that the nations of Europe are no longer able to conduct their own monetary policies. Instead, the European Central Bank, with the participation of all member countries, sets a single monetary policy for all of Europe. The central banks of the individual countries play a role similar to that of regional Federal Reserve Banks: they monitor local conditions but have no control over the money supply or interest rates. Critics of the move toward a common currency argue that the cost of losing national monetary policy is large. When a recession hits one country but not others in Europe, that country does not have the tools of monetary policy to combat the downturn. This argument is one reason some European nations, such as the United Kingdom and Sweden, have chosen not to give up their own currencies and adopt the euro.

The problems associated with giving up national monetary policy have become very apparent recently. From 2008 to 2013, several of the economies of southern Europe experienced pronounced downturns. The unemployment rate rose from 6.7 to 12.2 percent in Italy, 8.5 to 16.5 percent in Portugal, 11.3 to 26.1 percent in Spain, and 7.7 to 27.3 percent in Greece. By contrast, in Germany, the largest country using the euro, the unemployment rate fell from 7.5 to 5.3 percent during this period. Critics of the euro contend that if these southern European nations had their own currencies, rather than being part of the euro area with Germany, they could have pursued more expansionary monetary policy. Such a move would have weakened their currencies and made their exports less expensive on world markets; the increase in net exports would have helped maintain aggregate demand and soften the recession.

Why, according to euro critics, is monetary union a bad idea for Europe if it works well in the United States? These economists argue that the United States is different from Europe in two important ways. First, labor is more mobile among U.S. states than among European countries. This is in part because the United States has a common language and in part because most Americans are descended from immigrants, who have shown a willingness to move. Therefore, when a regional recession occurs, U.S. workers are more likely to move from high-unemployment states to low-unemployment states. Second, the United States has a strong central government that can use fiscal policy—such as the federal income tax—to redistribute resources among regions. Because Europe does not have these two advantages, it bears a larger cost when it adopts a single monetary policy.

Advocates of a common currency believe that the loss of national monetary policy is more than offset by other gains. With a single European currency, travelers and businesses worry less about exchange rates, and this encourages more international trade. A common currency may also have the political advantage of making Europeans feel more connected to one another. The twentieth century was marked by two world wars sparked by European discord. If a common currency makes the nations of Europe more harmonious, euro advocates argue, it benefits the entire world. ■

Speculative Attacks, Currency Boards, and Dollarization

Imagine you are a central banker of a small country. You and your fellow policymakers decide to fix your currency—let's call it the peso—against the U.S. dollar. From now on, one peso will sell for one dollar.

As we discussed earlier, you now have to stand ready to buy and sell pesos for a dollar each. The money supply will adjust automatically to make the equilibrium exchange rate equal your target. There is, however, one potential problem with this plan: you might run out of dollars. If people come to the central bank to sell large quantities of pesos, the central bank's dollar reserves might dwindle to zero. In this case, the central bank has no choice but to abandon the fixed exchange rate and let the peso depreciate.

This fact raises the possibility of a *speculative attack*—a change in investors' perceptions that makes the fixed exchange rate untenable. Suppose that, for no good reason, a rumor spreads that the central bank is going to abandon the exchange-rate peg. People would respond by rushing to the central bank to convert pesos into dollars before the pesos lose value. This rush would drain the central bank's reserves and could force the central bank to abandon the peg. In this case, the rumor would prove self-fulfilling.

To avoid this possibility, some economists argue that a fixed exchange rate should be supported by a *currency board*, such as that used by Argentina in the 1990s. A currency board is an arrangement by which the central bank holds enough foreign currency to back each unit of the domestic currency. In our example, the central bank would hold one U.S. dollar (or one dollar invested in a U.S. government bond) for every peso. No matter how many pesos turned up at the central bank to be exchanged, the central bank would never run out of dollars.

Once a central bank has adopted a currency board, it might consider the natural next step: abandoning the peso altogether and letting its country use the U.S. dollar. Such a plan is called *dollarization*. It happens on its own in high-inflation economies, where foreign currencies offer a more reliable store of value than the domestic currency. But it can also occur as a matter of public policy, as in Panama. If a country wants its currency to be irrevocably fixed to the dollar, the most reliable method is to make the dollar its official currency. The only loss from dollarization is the seigniorage revenue that a government gives up by relinquishing its control over the printing press. The U.S. government then gets the revenue generated by growth in the money supply.[5]

The Impossible Trinity

The analysis of exchange-rate regimes leads to a simple conclusion: you can't have it all. To be more precise, it is impossible for a nation to have free capital flows, a fixed exchange rate, and independent monetary policy. This fact, often called the **impossible trinity** (or sometimes the *trilemma of international finance*), is illustrated in Figure 13-12. A nation must choose one side of this triangle, giving up the institutional feature at the opposite corner.

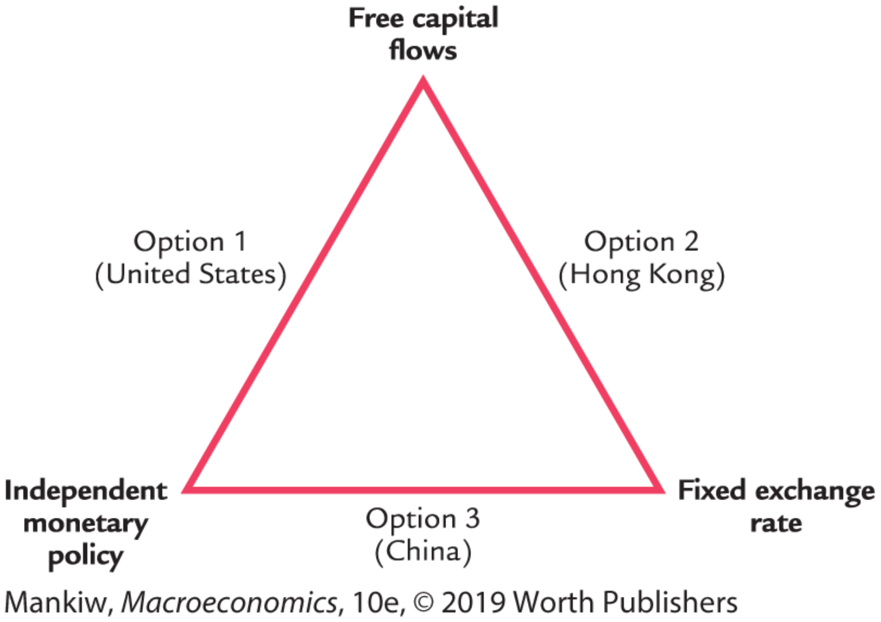


FIGURE 13-12 The Impossible Trinity It is impossible for a nation to have free capital flows, a fixed exchange rate, and independent monetary policy. A nation must choose one side of this triangle and give up the opposite corner.

The first option is to allow free flows of capital and to conduct an independent monetary policy, as the

United States has done in recent years. In this case, it is impossible to have a fixed exchange rate. Instead, the exchange rate must float to equilibrate the market for foreign-currency exchange.

The second option is to allow free flows of capital and to fix the exchange rate, as Hong Kong has done in recent years. In this case, the nation loses the ability to conduct an independent monetary policy. The money supply must adjust to keep the exchange rate at its predetermined level. In a sense, when a nation fixes its currency to that of another nation, it is adopting that other nation's monetary policy.

The third option is to restrict the international flow of capital in and out of the country, as China has done in recent years. In this case, the interest rate is no longer fixed by world interest rates but is determined by domestic forces, as in a closed economy. It is then possible to both fix the exchange rate and conduct an independent monetary policy.

History has shown that nations can, and do, choose different sides of the trinity. A nation's policymakers must ask themselves the following question: do they want to live with exchange-rate volatility (option 1), do they want to give up the use of monetary policy for purposes of domestic stabilization (option 2), or do they want to restrict their citizens from participating in world financial markets (option 3)? Every nation must make one of these choices.

CASE STUDY

The Chinese Currency Controversy

From 1995 to 2005 the Chinese currency, the yuan, was pegged to the dollar at an exchange rate of 8.28 yuan per U.S. dollar. In other words, China's central bank stood ready to buy and sell yuan at this price. This policy of fixing the exchange rate was combined with a policy of restricting international capital flows. Chinese citizens were not allowed to convert their savings into dollars or euros and invest abroad.

By the early 2000s, many observers believed that the yuan was significantly undervalued. They suggested that if the yuan were allowed to float, it would increase in value relative to the dollar. The evidence in favor of this hypothesis was that China was accumulating large dollar reserves in its efforts to maintain the fixed exchange rate. That is, China's central bank had to supply yuan and demand dollars in foreign-exchange markets to keep the yuan at the pegged level. If this intervention in the currency market ceased, the yuan would rise in value compared to the dollar.

The pegged yuan became a contentious political issue in the United States. U.S. producers that competed against Chinese imports complained that the undervalued yuan made Chinese goods cheaper, putting the U.S. producers at a disadvantage. (U.S. consumers benefited from inexpensive imports, but in the politics of international trade, producers often shout louder than consumers.) In response to these concerns, President George W. Bush called on China to let its currency float. Several senators proposed a more drastic step—a steep tariff on Chinese imports until China adjusted the value of its currency.

China no longer completely fixes the exchange rate. In July 2005 China announced a new policy: it would still intervene in foreign-exchange markets to prevent large and sudden movements in the exchange rate, but it would permit gradual changes. Moreover, it would judge the value of the yuan not just relative to the dollar but also relative to a broad basket of currencies. Over the next decade, the yuan appreciated by about 25 percent. China's critics, including President Donald Trump, at times still complain about that nation's intervention in foreign-exchange markets, but today China's exchange-rate policy is a less pressing issue on the international economic agenda than it was in the past. ■

13-6 From the Short Run to the Long Run: The Mundell–Fleming Model with a Changing Price Level

So far we have used the Mundell–Fleming model to study the small open economy in the short run when the price level is fixed. We now consider what happens when the price level changes. Doing so will show how the Mundell–Fleming model provides a theory of the aggregate demand curve in a small open economy. It will also show how this short-run model relates to the long-run model of the open economy we examined in Chapter 6.

Because we now want to consider changes in the price level, the nominal and real exchange rates in the economy will no longer be moving in tandem. Thus, we must distinguish between these two variables. The nominal exchange rate is *e* and the real exchange rate is ε, which equals eP/P^*, as you should recall from Chapter 6. We can write the Mundell–Fleming model as

$$\begin{aligned} Y &= C(Y-T)+I(r^*)+G+NX(\varepsilon) && IS^*, \\ M/P &= L(r^*,\,Y) && LM^*. \end{aligned}$$

These equations should be familiar by now. The first equation describes the IS^* curve, and the second describes the LM^* curve. Note that net exports depend on the real exchange rate.

Figure 13-13 shows what happens when the price level falls. Because a lower price level raises the level of real money balances, the LM^* curve shifts to the right, as in panel (a). The real exchange rate falls, and income rises. The aggregate demand curve summarizes this negative relationship between the price level and income, as shown in panel (b).

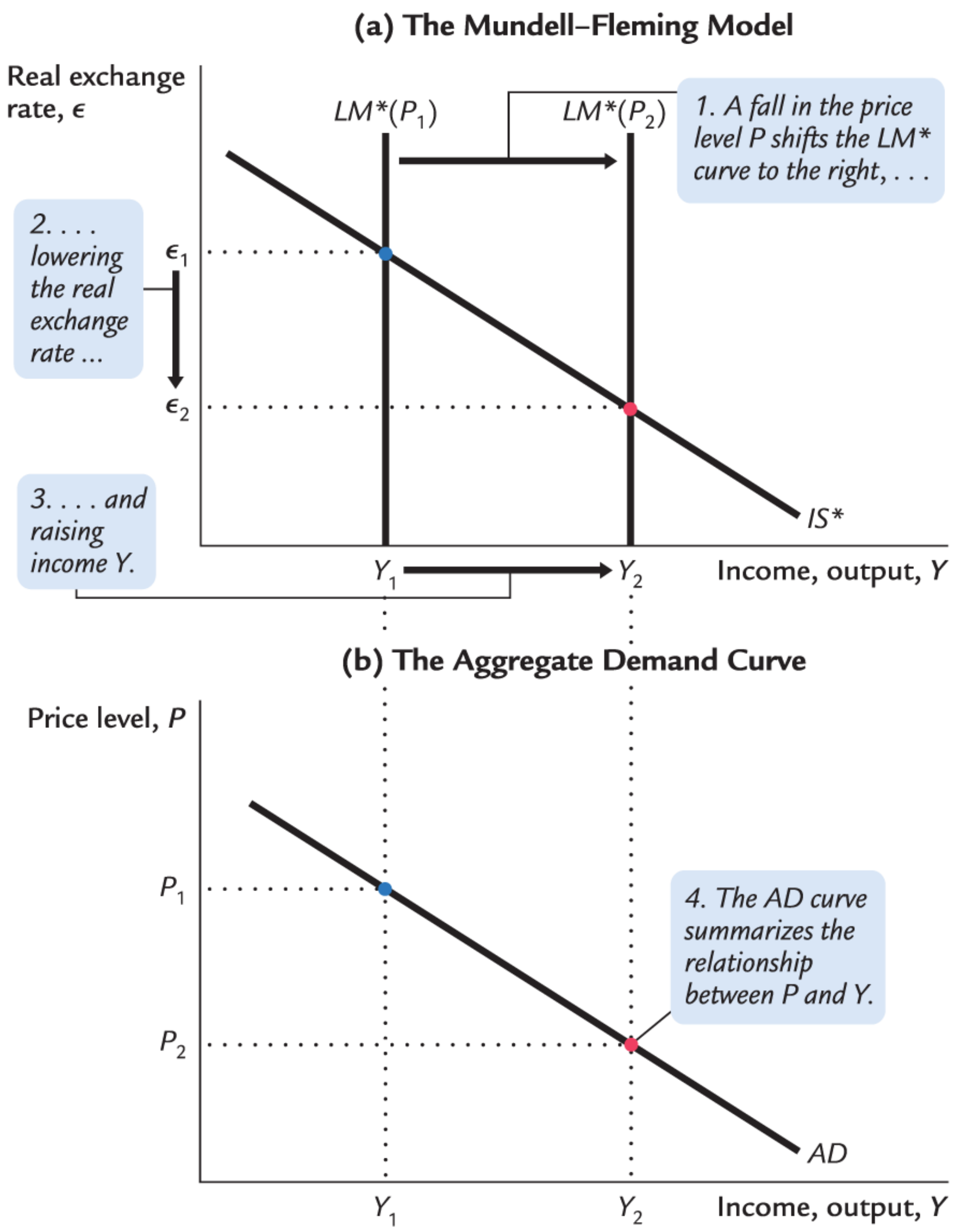
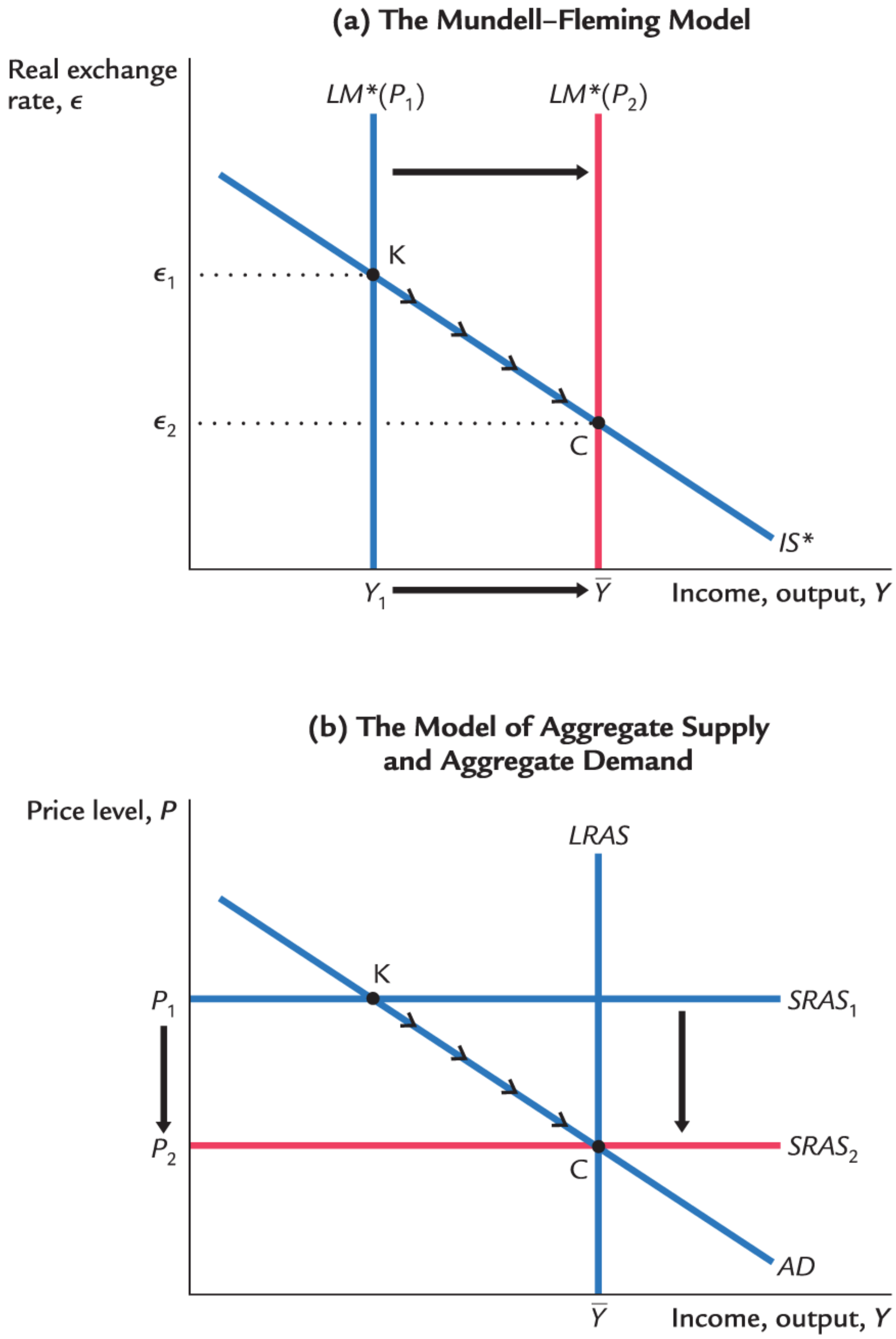


FIGURE 13-13 Mundell–Fleming as a Theory of Aggregate Demand Panel (a) shows that when the price level falls, the LM LM^* curve shifts to the right. As a result, equilibrium income rises. Panel (b) shows that this negative relationship between *P* and *Y* is summarized by the aggregate demand curve.

Thus, just as the *IS*–*LM* model explains the aggregate demand curve in a closed economy, the Mundell–Fleming model explains the aggregate demand curve for a small open economy. In both cases, the aggregate demand curve shows the set of equilibria in the goods and money markets that arise as the price level varies. And in both cases, anything that changes equilibrium income, other than a change in the price level, shifts the aggregate demand curve. Policies and events that raise income for a given price level shift the aggregate demand curve to the right; policies and events that lower income for a given price level shift the aggregate demand curve to the left.

We can use this diagram to show how the short-run model in this chapter is related to the long-run model in Chapter 6. Figure 13-14 shows the short-run and long-run equilibria. In both panels of the figure, point K describes the short-run equilibrium because it assumes a fixed price level. At this equilibrium, the demand for goods and services is too low to keep the economy producing at its natural level. Over time, low demand causes the price level to fall. The fall in the price level raises real money balances, shifting the *LM* curve to the right. The real exchange rate depreciates, so net exports rise. Eventually, the economy reaches point C, the long-run equilibrium. The speed of transition between the short-run and long-run equilibria depends on how quickly the price level adjusts to restore the economy to the natural level of output.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-14 The Short-Run and Long-Run Equilibria in a Small Open Economy Point K in both panels shows the equilibrium under the Keynesian assumption that the price level is fixed at P1. P_1. Point C in both panels shows the equilibrium under the classical assumption that the price level adjusts to maintain income at its natural level Y‾. $\bar{Y}$.

The levels of income at point K and point C are both of interest. Our central concern in this chapter has been how policy influences point K, the short-run equilibrium. In Chapter 6 we examined the determinants of point C, the long-run equilibrium. Whenever policymakers consider any change in policy, they need to consider both the short-run and long-run effects of their decision.

13-7 A Concluding Reminder

In this chapter we have examined how a small open economy works in the short run when prices are sticky. We have seen how monetary, fiscal, and trade policies influence income and the exchange rate, as well as how the behavior of the economy depends on whether the exchange rate is floating or fixed. In closing, it is worth repeating a lesson from Chapter 6. Many countries, including the United States, are neither closed economies nor small open economies: they lie somewhere in between.

A large open economy, such as that of the United States, combines the behavior of a closed economy and the behavior of a small open economy. When analyzing policies in a large open economy, we need to consider both the closed-economy logic of Chapter 12 and the open-economy logic developed in this chapter. The appendix to this chapter presents a model for a large open economy. The results are, as one might guess, a mixture of the two polar cases we have already examined.

To see how we can draw on the logic of both the closed and small open economies and apply these insights to the United States, consider how a monetary contraction affects the economy in the short run. In a closed economy, a monetary contraction raises the interest rate, lowers investment, and thus lowers aggregate income. In a small open economy with a floating exchange rate, a monetary contraction raises the exchange rate, lowers net exports, and thus lowers aggregate income. The interest rate is unaffected, however, because it is determined by world financial markets.

The U.S. economy has elements of both cases. Because the United States is large enough to affect the world interest rate and because capital is not perfectly mobile across countries, a monetary contraction raises the interest rate and depresses investment. At the same time, a monetary contraction also raises the value of the dollar, thereby depressing net exports. Hence, although the Mundell–Fleming model does not precisely describe an economy like that of the United States, it correctly predicts what happens to international variables such as the exchange rate, and it shows how international interactions alter the effects of monetary and fiscal policies.

APPENDIX

A Short-Run Model of the Large Open Economy

When analyzing policies in an economy such as that of the United States, we need to combine the closed-economy logic of the *IS*–*LM* model and the small-open-economy logic of the Mundell–Fleming model. This appendix presents a model for the intermediate case of a large open economy.

As we discussed in the appendix to Chapter 6, a large open economy differs from a small open economy because its interest rate is not fixed by world financial markets. In a large open economy, we must consider the relationship between the interest rate and the flow of capital abroad. The net capital outflow is the amount that domestic investors lend abroad minus the amount that foreign investors lend here. As the domestic interest rate falls, domestic investors find foreign lending more attractive, and foreign investors find lending here less attractive. Thus, the net capital outflow is negatively related to the interest rate. Here we add this relationship to our short-run model of national income.

The three equations of the model are

Y=C(Y−T)+I(r)+G+NX(e),M/P=L(r, Y),NX(e)=CF(r).

$$\begin{aligned} Y &= C(Y-T)+I(r)+G+NX(e), \\ M/P &= L(r,\ Y), \\ NX(e) &= CF(r). \end{aligned}$$

The first two equations are the same as those used in the Mundell–Fleming model of this chapter. The third equation, taken from the appendix to Chapter 6, states that the trade balance *NX* equals the net capital outflow *CF*, which in turn depends on the domestic interest rate.

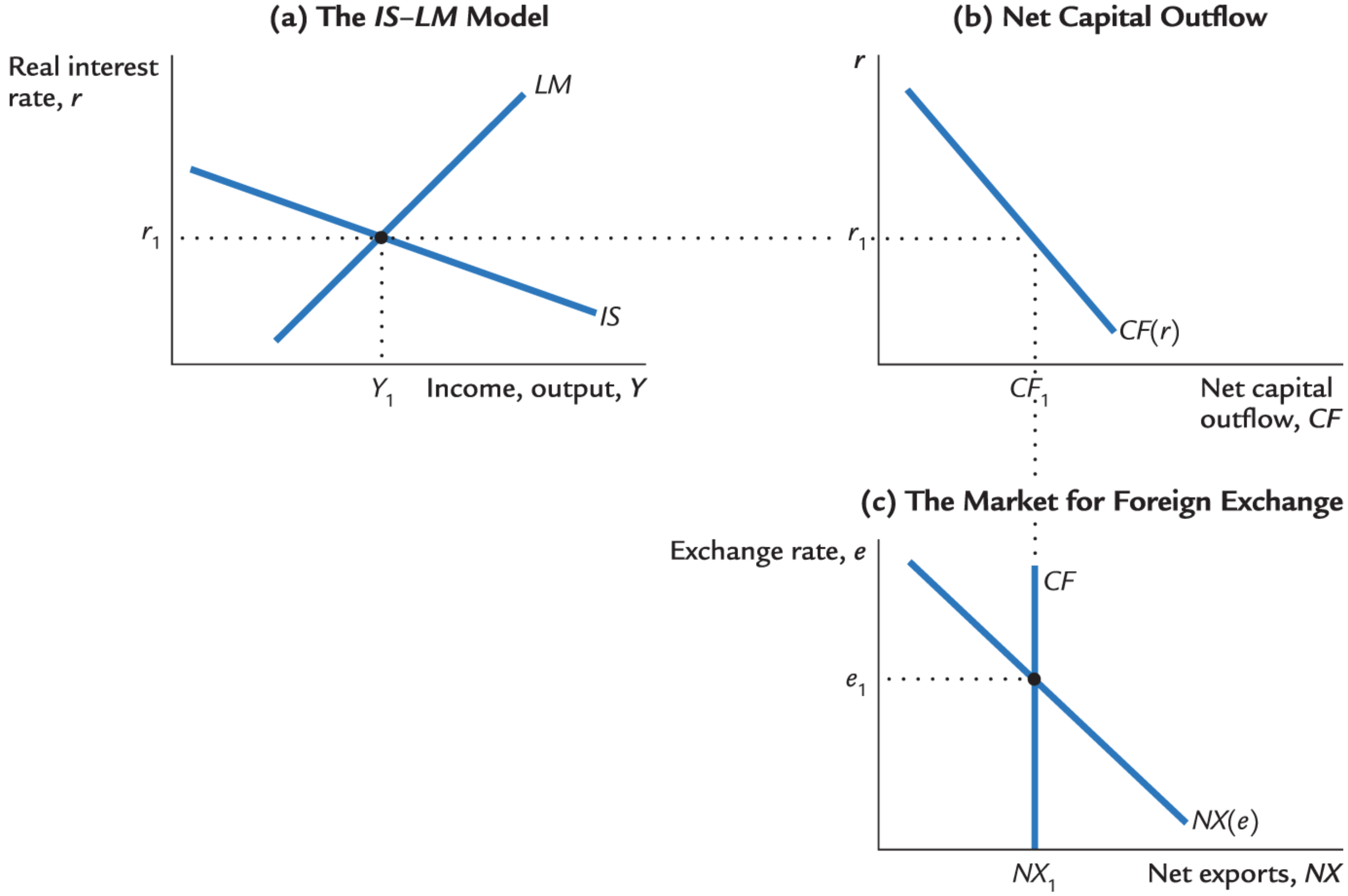
To see what this model implies, substitute the third equation into the first, so the model becomes

Y=C(Y−T)+I(r)+G+CF(r)IS,M/P=L(r, Y)LM.

$$\begin{aligned} Y &= C(Y-T)+I(r)+G+CF(r) \qquad IS, \\ M/P &= L(r,\ Y) \qquad LM. \end{aligned}$$

These two equations are much like the two equations of the closed-economy *IS*–*LM* model. The only difference is that expenditure now depends on the interest rate for two reasons. As before, a higher interest rate reduces investment. But now a higher interest rate also reduces the net capital outflow and thus lowers net exports.

To analyze this model, we can use the three graphs in Figure 13-15. Panel (a) shows the *IS*–*LM* diagram. As in the closed-economy model in Chapters 11 and 12, the interest rate *r* is on the vertical axis, and income *Y* is on the horizontal axis. The *IS* and *LM* curves together determine equilibrium income and the equilibrium interest rate.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-15 A Short-Run Model of a Large Open Economy Panel (a) shows that the *IS* and *LM* curves determine the interest rate r_1 and income Y_1. Panel (b) shows that r_1 determines the net capital outflow CF_1. Panel (c) shows that CF_1 and the net-exports schedule determine the exchange rate e_1.

The new net-capital-outflow term in the *IS* equation, *CF*(*r*), makes this *IS* curve flatter than it would be in a closed economy. The more responsive international capital flows are to the interest rate, the flatter the *IS* curve is. You might recall from the Chapter 6 appendix that the small open economy represents the extreme case in which the net capital outflow is infinitely elastic at the world interest rate. In this extreme case, the *IS* curve is completely flat. Hence, a small open economy would be depicted in this figure with a horizontal *IS* curve.

Panels (b) and (c) show how the equilibrium from the *IS–LM* model determines the net capital outflow, the trade balance, and the exchange rate. In panel (b) we see that the interest rate determines the net capital outflow. This curve slopes downward because a higher interest rate discourages domestic investors from lending abroad and encourages foreign investors to lend here, thereby reducing the net capital outflow. In panel (c) we see that the exchange rate adjusts to ensure that net exports of goods and services equal the net capital outflow.

Now let's use this model to examine the impact of various policies. We assume that the economy has a floating exchange rate because this assumption is correct for most large open economies such as that of the United States.

Fiscal Policy

Figure 13-16 examines the impact of a fiscal expansion. An increase in government purchases or a cut in taxes shifts the *IS* curve to the right. As panel (a) illustrates, this shift in the *IS* curve leads to an increase in income and an increase in the interest rate. These two effects are similar to those in a closed economy.

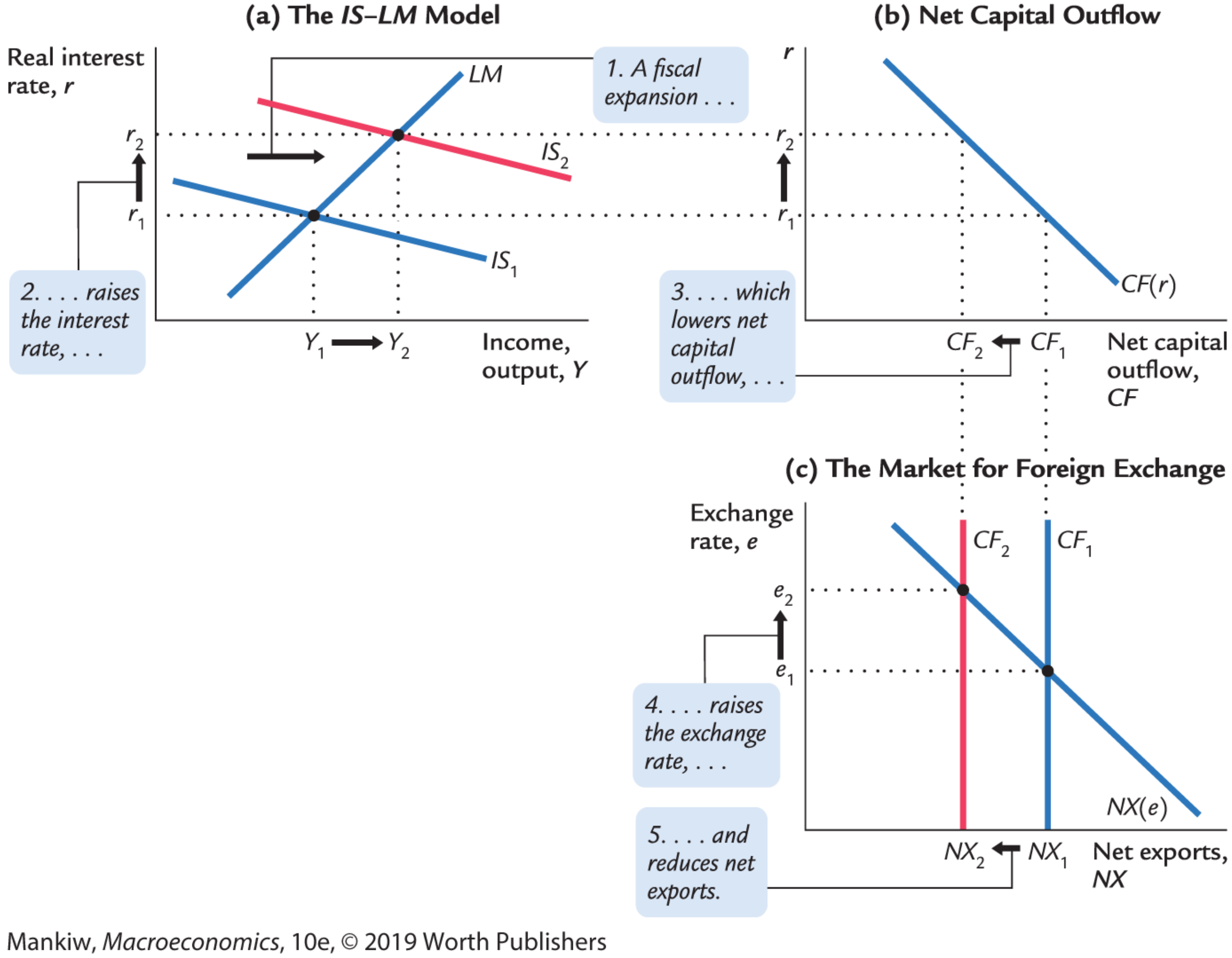


FIGURE 13-16 A Fiscal Expansion in a Large Open Economy Panel (a) shows that a fiscal expansion shifts the *IS*

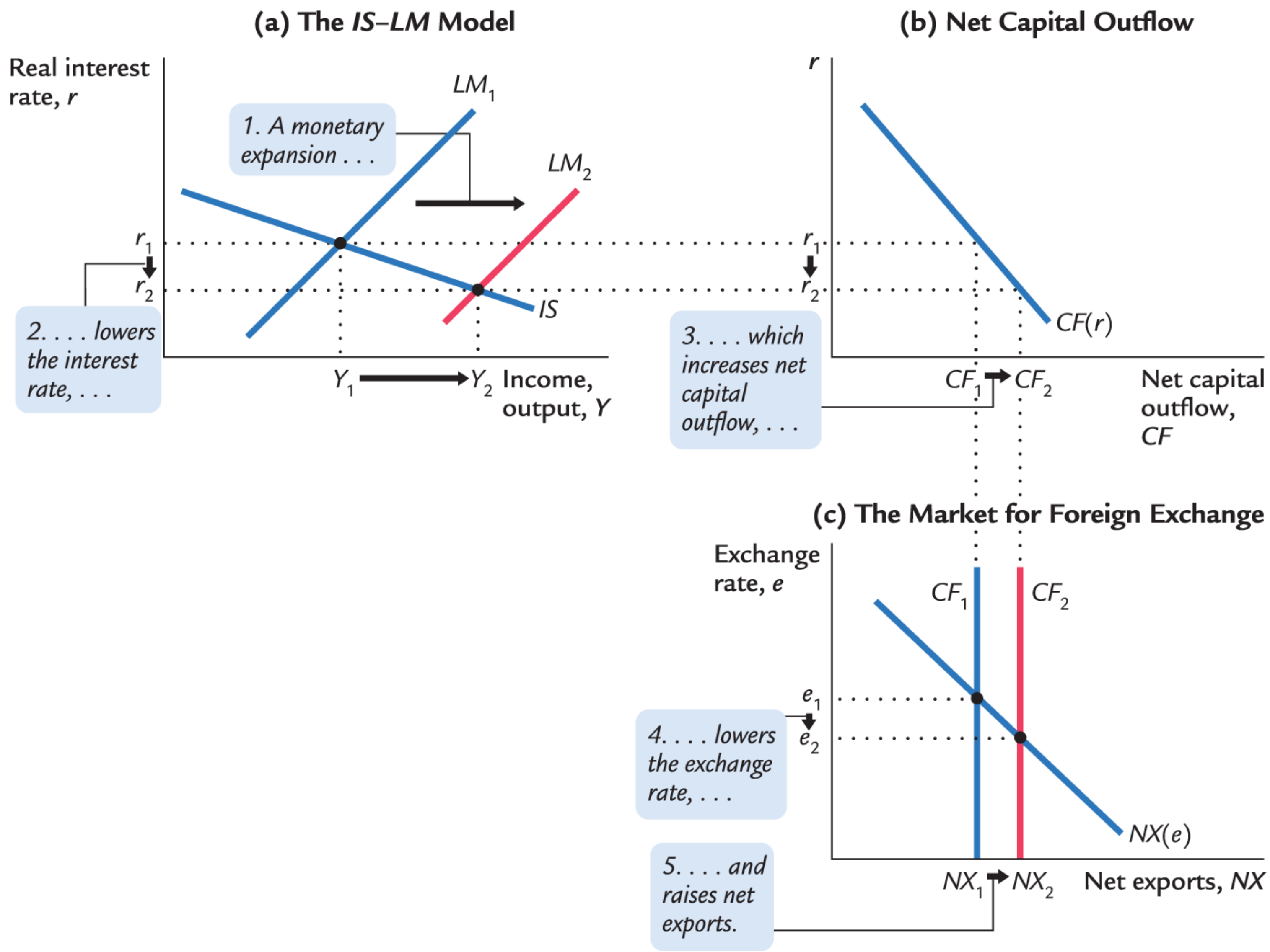
curve to the right. Income rises from Y_1 to Y_2, and the interest rate rises from r_1 to r_2. Panel (b) shows that the increase in the interest rate causes the net capital outflow to fall from CF_1 to CF_2. Panel (c) shows that the fall in the net capital outflow reduces the net supply of dollars, causing the exchange rate to rise from e_1 to e_2.

Yet in the large open economy the higher interest rate reduces the net capital outflow, as in panel (b). The fall in the net capital outflow reduces the supply of dollars in the market for foreign-currency exchange. The exchange rate appreciates, as in panel (c). Because domestic goods become more expensive relative to foreign goods, net exports fall.

Figure 13-16 shows that a fiscal expansion raises income in the large open economy, unlike in a small open economy under a floating exchange rate. The impact on income, however, is smaller than in a closed economy. In a closed economy, the expansionary impact of fiscal policy is partially offset by the crowding out of investment: as the interest rate rises, investment falls, reducing the fiscal-policy multipliers. In a large open economy, there is yet another offsetting factor: as the interest rate rises, the net capital outflow falls, the currency appreciates in the foreign-exchange market, and net exports fall. This reduces the fiscal-policy multiplier even further. (In the figure, this additional channel is manifested by the flatter *IS* curve mentioned earlier: for any given rightward shift in the *IS* curve, a flatter curve implies a smaller expansion in income.) Together these effects are not large enough to make fiscal policy powerless, as it is in a small open economy, but they do reduce the impact of fiscal policy.

Monetary Policy

Figure 13-17 examines the effect of a monetary expansion. An increase in the money supply shifts the *LM* curve to the right, as in panel (a). Income rises, and the interest rate falls. Once again, these effects are similar to those in a closed economy.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 13-17 A Monetary Expansion in a Large Open Economy Panel (a) shows that a monetary expansion shifts the *LM* curve to the right. Income rises from Y_1 to Y_2, and the interest rate falls from r_1 to r_2. Panel (b) shows that the decrease in the interest rate causes the net capital outflow to increase from CF_1 to CF_2. Panel (c) shows that the increase in the net capital outflow raises the net supply of dollars, causing the exchange rate to fall from e_1 to e_2.

Yet, as panel (b) shows, the lower interest rate leads to a higher net capital outflow. The increase in *CF* raises the supply of dollars in the market for foreign-currency exchange. The exchange rate falls, as in panel (c). As domestic goods become cheaper relative to foreign goods, net exports rise.

We can now see that the monetary transmission mechanism works through two channels in a large open economy. As in a closed economy, a monetary expansion lowers the interest rate, stimulating investment. As in a small open economy, a monetary expansion causes the currency to depreciate in the market for foreign-currency exchange, stimulating net exports. Both effects result in higher income. Indeed, because the *IS* curve is flatter here than it is in a closed economy, any given shift in the *LM* curve will have a larger impact on income.

A Rule of Thumb

This model of the large open economy describes well the U.S. economy today. Yet it is somewhat more complicated and cumbersome than the model of the closed economy we studied in Chapters 11 and 12 and the model of the small open economy we developed in this chapter. Fortunately, there is a useful rule of thumb to help you determine how policies influence a large open economy without remembering all the details of the model: *The large open economy is an average of the closed economy and the small open economy. To find how any policy will affect any variable, find the answer in the two extreme cases and take an average*.

For example, how does a monetary contraction affect the interest rate and investment in the short run? In a closed economy, the interest rate rises, and investment falls. In a small open economy, neither the interest rate nor investment changes. The effect in the large open economy is an average of these two cases: a monetary contraction raises the interest rate and reduces investment—but only somewhat. The fall in the net capital outflow mitigates the rise in the interest rate and the fall in investment that would occur in a closed economy. But unlike in a small open economy, the international flow of capital is not so strong as to fully negate these effects.

This rule of thumb makes the simple models all the more valuable. Although they do not describe perfectly the world in which we live, they do provide a useful guide to the effects of economic policy.

MORE PROBLEMS AND APPLICATIONS

1. Imagine that you run the central bank in a large open economy with a floating exchange rate. Your goal is to stabilize income, and you adjust the money supply accordingly. Under your policy, what happens to the money supply, the interest rate, the exchange rate, and the trade balance in response to each of the following shocks?
    a. The government raises taxes to reduce the budget deficit.
    b. The government restricts the import of foreign cars.

2. Over the past several decades, the economies of the world have become more financially integrated. That is, investors in all nations have become more willing and able to take advantage of financial opportunities abroad. Consider how this development affects the ability of monetary policy to influence the economy.
    a. If investors become more willing and able to substitute foreign and domestic assets, what happens to the slope of the *CF* function?
    b. If the *CF* function changes in this way, what happens to the slope of the *IS* curve?
    c. How does this change in the *IS* curve affect the Fed's ability to control the interest rate?
    d. How does this change in the *IS* curve affect the Fed's ability to control national income?

3. Suppose policymakers in a large open economy want to raise investment without changing income or the exchange rate.
    a. Is there any combination of domestic monetary and fiscal policies that would achieve this goal?

b. Is there any combination of domestic monetary, fiscal, and trade policies that would achieve this goal?
c. Is there any combination of monetary and fiscal policies at home and abroad that would achieve this goal?

4. This appendix considers the case of a large open economy with a floating exchange rate. Now suppose that a large open economy has a fixed exchange rate. That is, the central bank announces a target for the exchange rate and commits itself to adjusting the money supply to ensure that the equilibrium exchange rate equals the target.
   a. Describe what happens to income, the interest rate, and the trade balance in response to a fiscal expansion, such as an increase in government purchases. Compare your answer to the case of a small open economy with a fixed exchange rate.
   b. Describe what happens to income, the interest rate, and the trade balance if the central bank expands the money supply by buying bonds from the public. Compare your answer to the case of a small open economy with a fixed exchange rate.

CHAPTER 14

Aggregate Supply and the Short-Run Tradeoff Between Inflation and Unemployment

> Probably the single most important macroeconomic relationship is the Phillips curve.
>
> —George Akerlof

> There is always a temporary tradeoff between inflation and unemployment; there is no permanent tradeoff. The temporary tradeoff comes not from inflation per se, but from unanticipated inflation, which generally means, from a rising rate of inflation.
>
> —Milton Friedman

Most economists analyze short-run fluctuations in national income and the price level using the model of aggregate demand and aggregate supply. In the previous three chapters, we examined aggregate demand in some detail. The *IS*–*LM* model (and its open-economy cousin the Mundell–Fleming model) shows how changes in monetary and fiscal policy and shocks to the money and goods markets shift the aggregate demand curve. In this chapter, we consider what determines the position and slope of the aggregate supply curve.

When introducing the aggregate supply curve in Chapter 10, we established that aggregate supply behaves differently in the short run than in the long run. In the long run, prices are flexible, and the aggregate supply curve is vertical. When the aggregate supply curve is vertical, shifts in the aggregate demand curve affect the price level, but the output of the economy remains at its natural level. By contrast, in the short run, prices are sticky, and the aggregate supply curve is not vertical. In this case, shifts in aggregate demand cause fluctuations in output. Chapter 10 offered a simplified view of price stickiness: the short-run aggregate supply curve was a horizontal line, representing the extreme case in which all prices are fixed. Our task now is to refine this understanding of short-run aggregate supply to better reflect the real world, in which some prices are sticky while others are not.

After examining the theory of the short-run aggregate supply curve, we establish a key implication. We show that this curve implies a tradeoff between two measures of economic performance—inflation and unemployment. This tradeoff, called the *Phillips curve*, tells us that to reduce inflation, policymakers must temporarily raise unemployment, and to reduce unemployment, they must accept higher inflation. As Milton Friedman’s quotation at the beginning of the chapter suggests, the tradeoff between inflation and unemployment is only temporary. One goal of this chapter is to explain why policymakers face such a tradeoff in the short run and why they do not face it in the long run.

14-1 The Basic Theory of Aggregate Supply

When physics classes study balls rolling down inclined planes, they begin by assuming away the existence of friction. This simplification is a good starting point, but no engineer would ever take the no-friction assumption as a literal description of how the world works. Similarly, this book began with classical macroeconomic theory, but it would be a mistake to assume that this model is always true. Our job now is to look more deeply into the “frictions” of macroeconomics.

We do this by examining two models of aggregate supply. In both models, some market imperfection (that is, some type of friction) causes the output of the economy to deviate from its natural level. As a result, the short-run aggregate supply curve is upward sloping rather than vertical, and shifts in the aggregate demand curve cause output to fluctuate. These temporary deviations of output from its natural level represent the booms and busts of the business cycle.

Each of the two models takes us down a different theoretical route, but both routes end up in the same place. That final destination is a short-run aggregate supply equation of the form

$$Y = \overline{Y} + \alpha(P - EP),\ \alpha > 0,$$

where *Y* is output, $\overline{Y}$ is the natural level of output, *P* is the price level, and *EP* is the expected price level. This equation states that output deviates from its natural level when the price level deviates from the expected price level. The parameter α indicates how much output responds to unexpected changes in the price level; $1/\alpha$ is the slope of the aggregate supply curve.

The two models tell a different story about what lies behind this short-run aggregate supply equation. In other words, each model highlights a particular reason unexpected movements in the price level are associated with fluctuations in aggregate output.

The Sticky-Price Model

The most widely accepted explanation for the upward-sloping short-run aggregate supply curve is called the **sticky-price model**. This model emphasizes that firms do not instantly adjust the prices they charge in response to changes in demand. Sometimes prices are set by long-term contracts between firms and customers. Even without formal agreements, firms may hold prices steady to avoid annoying their regular customers with frequent price changes. Some prices are sticky because of the way certain markets are structured: once a firm has printed and distributed its catalog or price list, it is costly to alter prices. And sometimes sticky prices reflect sticky wages: firms base their prices on the costs of production, and wages may depend on social norms and notions of fairness that evolve only slowly over time.

There are various ways to formalize the idea of sticky prices as the basis for an upward-sloping aggregate supply curve. Here we examine an especially simple model. We first consider the pricing decisions of individual firms and then add together the decisions of many firms to explain the behavior of the economy as a whole. To develop the model, we depart from the assumption of perfect competition, which we have used since Chapter 3. Perfectly competitive firms are price-takers, not price-setters. When considering how firms set prices, we now assume that these firms have some market power over the prices they charge.

Consider the pricing decision facing a typical firm. The firm's desired price *p* depends on two macroeconomic variables:

- The overall level of prices *P*. A higher price level implies that the firm's costs are higher. Hence, the higher the price level, the more the firm would like to charge for its product.
- The level of aggregate income *Y*. Higher income raises the demand for the firm's product. Because marginal cost increases at higher levels of production, the greater the demand, the higher the firm's desired price.

We write the firm's desired price as

$$p = P + a\left(Y - \overline{Y}\right).$$

This equation says that the desired price *p* depends on the overall level of prices *P* and on aggregate output relative to the natural level $Y - \overline{Y}$. The parameter *a* (which is greater than zero) measures how much the firm's desired price responds to aggregate output.[1]

Now assume that there are two types of firms. Some have flexible prices: they always set their prices according to this equation. Others have sticky prices: they announce their prices in advance, based on what they expect economic conditions to be. Firms with sticky prices set prices according to

$$p = EP + a\left(EY - E\overline{Y}\right),$$

where, as before, *E* represents the expected value of a variable. For simplicity, assume that these firms expect output to be at its natural level, so that the last term, $a\left(EY - E\overline{Y}\right)$, is zero. Then these firms set the price

$$p = EP.$$

That is, firms with sticky prices set their prices based on what they expect other firms to charge.

We can use the pricing rules of the two groups of firms to derive the aggregate supply equation. To do this, we find the economy's overall price level, which is the weighted average of the prices set by the two groups. If *s* is the fraction of firms with sticky prices and $1 - s$ is the fraction with flexible prices, then the price level is

$$P = sEP + (1 - s)\left[P + a\left(Y - \overline{Y}\right)\right].$$

The first term is the price of the sticky-price firms weighted by their fraction in the economy; the second term is the price of the flexible-price firms weighted by their fraction. Now subtract $(1 - s)P$ from both sides of this equation to obtain

$$sP = sEP + (1 - s)\left[a\left(Y - \overline{Y}\right)\right].$$

Divide both sides by *s* to solve for the price level:

$$P = EP + \left[(1 - s)a/s\right]\left(Y - \overline{Y}\right).$$

The two terms in this equation are explained as follows:

- When firms expect a high price level, they expect high costs. Firms that fix prices in advance set their prices high. These high prices cause the other firms to set high prices also. Hence, a high expected price level *EP* leads to a high actual price level *P*. This effect does not depend on the fraction of firms with sticky prices.
- When output is high, the demand for goods is high. Firms with flexible prices set their prices high, which leads to a high price level. The effect of output on the price level depends on the fraction of firms with sticky prices. The more firms there are with sticky prices, the less the price level responds to the level of economic activity.

Hence, the price level depends on the expected price level and on output.

Algebraic rearrangement puts this aggregate pricing equation into a more familiar form:

$$Y = \bar{Y} + \alpha (P - EP),$$

where $\alpha = s / [(1 - s) a]$. The sticky-price model says that the deviation of output from the natural level is positively associated with the deviation of the price level from the expected price level.[2]

An Alternative Theory: The Imperfect-Information Model

Another explanation for the upward slope of the short-run aggregate supply curve is called the **imperfect-information model**. Unlike the sticky-price model, this model assumes that markets clear—that is, all prices are free to adjust to balance supply and demand. In this model, the short-run and long-run aggregate supply curves differ because of temporary misperceptions about prices.

The imperfect-information model assumes that each supplier in the economy produces a single good and consumes many goods. Because the number of goods is so large, suppliers cannot always observe all prices. They monitor closely the prices of what they produce but less closely the prices of all the goods they consume. Because of imperfect information, they sometimes confuse changes in the price level with changes in relative prices. This confusion influences decisions about how much to supply, and it leads to a positive relationship between the price level and output in the short run.

Consider the decision facing a single supplier—an asparagus farmer, for instance. Because the farmer earns income from selling asparagus and uses this income to buy goods and services, the amount of asparagus he chooses to produce depends on the price of asparagus relative to the prices of other goods and services. If the relative price of asparagus is high, the farmer is motivated to work hard and produce more asparagus because the reward is great. If the relative price of asparagus is low, he would rather enjoy leisure and produce less asparagus.

Unfortunately, when making his production decision, the farmer does not know the relative price of asparagus. As an asparagus producer, he monitors the asparagus market closely and always knows the nominal price of asparagus. But he does not know the prices of all the other goods in the economy. He must, therefore, estimate the relative price of asparagus using the nominal price of asparagus and his expectation of the overall price level.

Consider how the farmer responds if all prices in the economy, including the price of asparagus, increase. One possibility is that he expected this change in prices. When he observes an increase in the price of asparagus, his estimate of its relative price is unchanged. He does not work any harder.

The other possibility is that the farmer did not expect the price level to increase (or to increase by this much). When he observes the increase in the price of asparagus, he is not sure whether other prices have risen (in which case the relative price of asparagus is unchanged) or whether only the price of asparagus has risen (in which case its relative price is higher). The rational inference is that some of each has happened. In other words, the farmer infers from the increase in the nominal price of asparagus that its relative price has risen somewhat. He works harder and produces more.

Our asparagus farmer is not unique. His decisions are similar to those of his neighbors, who produce broccoli, cauliflower, dill, eggplant, . . . , and zucchini. When the price level rises unexpectedly, all suppliers in the economy observe increases in the prices of the goods they produce. They all infer, rationally but mistakenly, that the relative prices of the goods they produce have risen. They work harder and produce more.

To sum up, the imperfect-information model says that when actual prices exceed expected prices, suppliers raise their output. The model implies an aggregate supply curve with the familiar form

$$Y = \overline{Y} + \alpha (P - EP).$$

Output deviates from its natural level when the price level deviates from the expected price level.

The imperfect-information story described above is the version developed originally by Nobel Prize–winning economist Robert Lucas in the 1970s. Recent work on imperfect-information models of aggregate supply has taken a somewhat different approach. Rather than emphasize confusion about relative prices and the absolute price level, as Lucas did, this new work stresses the speed at which information about the economy is incorporated into decisions. In this case, the friction that causes the short-run aggregate supply curve to slope upward is not the limited availability of information but is, instead, the limited ability of people to absorb and process information that is widely available. This information-processing constraint causes price-setters to respond slowly to economic news. The resulting equation for short-run aggregate supply is similar to those from the two models we have seen, even though the microeconomic foundations are somewhat different.[3]

CASE STUDY

International Differences in the Aggregate Supply Curve

All countries experience economic fluctuations, but these fluctuations are not the same everywhere. International differences are intriguing puzzles in themselves, and they often provide a way to test alternative theories. Examining international differences has been especially fruitful in research on aggregate supply.

When Robert Lucas proposed the imperfect-information model, he derived a surprising interaction between aggregate demand and aggregate supply: according to his model, the slope of the aggregate supply curve should depend on the volatility of aggregate demand. In countries where aggregate demand fluctuates widely, the aggregate price level fluctuates widely as well. Because most movements in prices in these countries do not represent movements in relative prices, suppliers should have learned not to respond much to unexpected changes in the price level. Therefore, the aggregate supply curve should be relatively steep (that is, α will be small). Conversely, in countries where aggregate demand is relatively stable, suppliers should have learned that most price changes are relative price changes. Accordingly, in these countries, suppliers should be more responsive to unexpected price changes, making the aggregate supply curve relatively flat (that is, α will be large).

Lucas tested this prediction using international data on output and prices. He found that changes in aggregate demand have the biggest effect on output in countries where aggregate demand and prices are most stable. Lucas concluded that the evidence supports the imperfect-information model.[4]

The sticky-price model also makes predictions about the slope of the short-run aggregate supply curve. In particular, it predicts that the average rate of inflation should influence the slope of the short-run aggregate supply curve. When the average rate of inflation is high, it is very costly for firms to keep prices fixed for long intervals. Thus, firms adjust prices more frequently. More frequent price adjustment in turn allows the price level to respond more quickly to shocks to aggregate demand. Hence, a high rate of inflation should make the short-run aggregate supply curve steeper.

International data support this prediction of the sticky-price model. In countries with low average inflation, the short-run aggregate supply curve is relatively flat: fluctuations in aggregate demand have large effects on output and are only slowly reflected in prices. High-inflation countries have steep short-run aggregate supply curves. In other words, high inflation appears to erode the frictions that cause prices to be sticky.[5]

Note that the sticky-price model can also explain Lucas's finding that countries with variable aggregate demand have steep aggregate supply curves. If the price level is highly variable, few firms will commit to prices in advance (s will be small). Hence, the aggregate supply curve will be steep (α will be small). ■

Implications

We have seen two models of aggregate supply and the market imperfection that each uses to explain why the short-run aggregate supply curve slopes upward. One model assumes the prices of some goods are sticky; the second assumes information about prices is imperfect. Keep in mind that these models are not incompatible with each other. We need not accept one model and reject the other. The world may contain both of these market imperfections, as well as some others, and all of them may contribute to the behavior of short-run aggregate supply.

The two models of aggregate supply differ in their assumptions and emphases, but their implications for output are similar. Both lead to the equation

$$Y = \overline{Y} + \alpha(P - EP).$$

This equation states that deviations of output from its natural level are related to deviations of the price level from the expected price level. *If the price level is higher than the expected price level, output exceeds its natural level. If the price level is lower than the expected price level, output falls short of its natural level.* Figure 14-1 graphs this equation. Note that the short-run aggregate supply curve is drawn for a given expectation *EP* and that a change in *EP* would shift the curve.

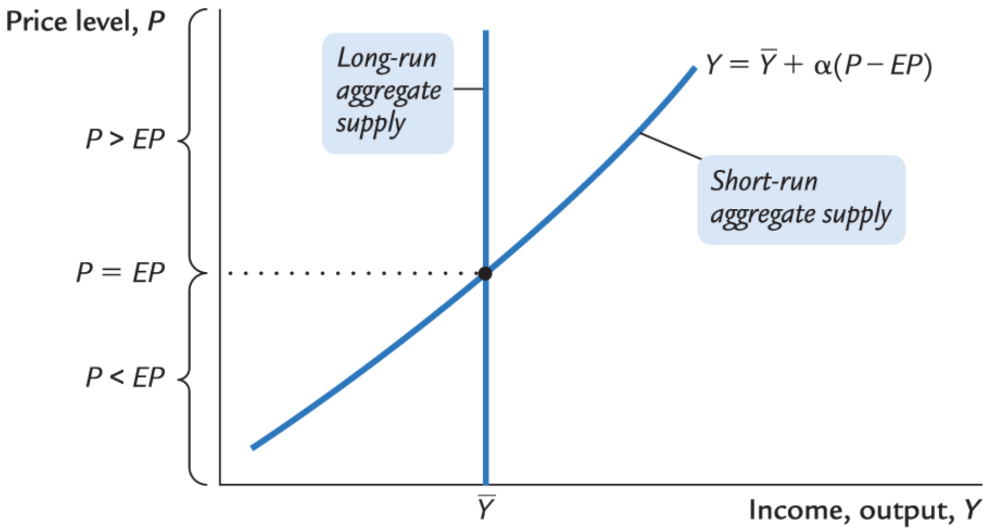
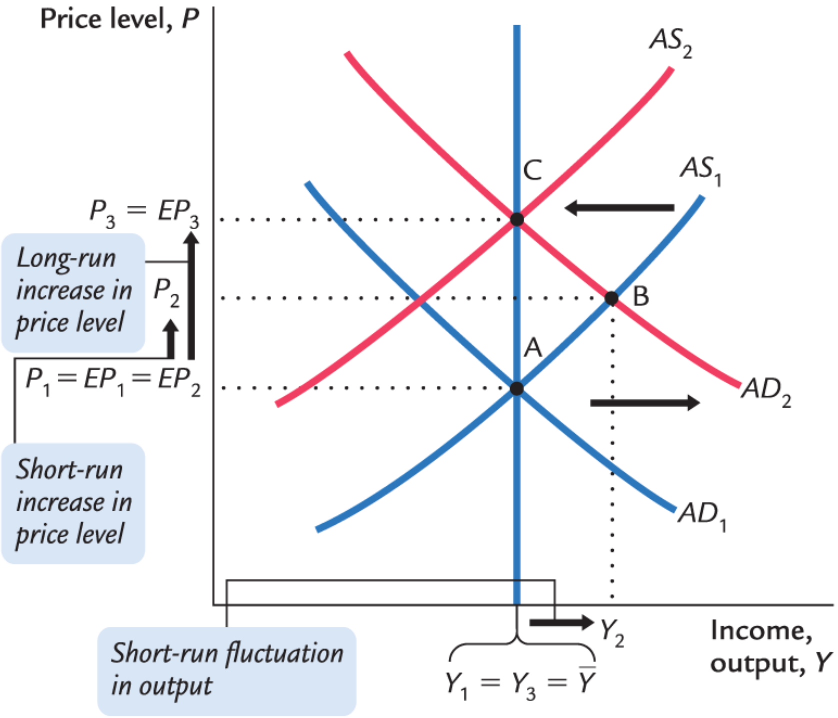


FIGURE 14-1 The Short-Run Aggregate Supply Curve Output deviates from its natural level $\overline{Y}$ if the price level *P* deviates from the expected price level *EP*.

Now that we have a better understanding of aggregate supply, let's put aggregate supply and aggregate demand back together. Figure 14-2 uses our aggregate supply equation to show how the economy responds to an unexpected increase in aggregate demand attributable, say, to an unexpected monetary expansion. In the short run, the equilibrium moves from point A to point B. The increase in aggregate demand raises the actual price level from P_1 to P_2. Because people did not expect this increase in the price level, the expected price level remains at EP_2, and output rises from Y_1 to Y_2, which is above the natural level $\overline{Y}$. Thus, the unexpected expansion in aggregate demand causes the economy to boom.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 14-2 How Shifts in Aggregate Demand Lead to Short-Run Fluctuations The economy begins in a long-run equilibrium, point A. When aggregate demand increases unexpectedly, the price level rises from P1P_1 to P2P_2. Because the price level P2P_2 is above the expected price level EP2,EP_2, output rises temporarily above the natural level, as the economy moves along the short-run aggregate supply curve from point A to point B. In the long run, the expected price level rises to EP3,EP_3, causing the short-run aggregate supply curve to shift upward. The economy returns to a new long-run equilibrium, point C, where output is back at its natural level.

But the boom does not last forever. In the long run, the expected price level rises to catch up with reality, causing the short-run aggregate supply curve to shift upward. As the expected price level rises from EP2EP_2 to EP3,EP_3, the equilibrium of the economy moves from point B to point C. The actual price level rises from P2P_2 to P3,P_3, and output falls from Y2Y_2 to Y3.Y_3. In other words, the economy returns to the natural level of output in the long run but at a much higher price level.

This analysis demonstrates a principle that holds for both models of aggregate supply: long-run monetary neutrality and short-run monetary *non*neutrality are compatible. Short-run nonneutrality is represented here by the movement from point A to point B, and long-run monetary neutrality is represented by the movement from point A to point C. We reconcile the short-run and long-run effects of money by emphasizing the adjustment of expectations about the price level.

14-2 Inflation, Unemployment, and the Phillips Curve

Two goals of economic policymakers are low inflation and low unemployment, but these goals can conflict. Suppose, for instance, policymakers were to use monetary or fiscal policy to expand aggregate demand. This policy would move the economy along the short-run aggregate supply curve to a point of higher output and a higher price level. (Figure 14-2 shows this as the change from point A to point B.) Higher output means lower unemployment because firms employ more workers when they produce more. A higher price level, given the previous year's price level, means higher inflation. Thus, when policymakers move the economy up along the short-run aggregate supply curve, they reduce unemployment and raise inflation. Conversely, when they contract aggregate demand and move the economy down the short-run aggregate supply curve, unemployment rises and inflation falls.

This tradeoff between inflation and unemployment, called the *Phillips curve*, is our topic in this section. As we have just seen (and will derive more formally in a moment), the Phillips curve reflects the short-run aggregate supply curve: as policymakers move the economy along the short-run aggregate supply curve, unemployment and inflation move in opposite directions. The Phillips curve is useful for expressing aggregate supply because inflation and unemployment are such important measures of economic performance.

Deriving the Phillips Curve from the Aggregate Supply Curve

The **Phillips curve** in its modern form states that the inflation rate depends on three forces:

- expected inflation;
- the deviation of unemployment from its natural rate, called *cyclical unemployment*; and
- supply shocks.

These three forces are expressed in the following equation:

π=Eπ−β(u−un)+vInflation=Expected Inflation−(β×Cyclical Unemployment)+Supply Shock

$$\begin{array}{ccccccccc} \pi & = & E\pi & - & \beta(u-u^n) & + & v \\ \text{Inflation} & = & \text{Expected Inflation} & - & (\beta \times \text{Cyclical Unemployment}) & + & \text{Supply Shock} \end{array}$$

where β is a parameter that measures the response of inflation to cyclical unemployment. Notice that there is a minus sign before the cyclical unemployment term: other things equal, higher unemployment is associated with lower inflation.

Where does this equation for the Phillips curve come from? We can derive it from our equation for aggregate supply. To see how, write the aggregate supply equation as

$$P = EP + (1/\alpha)\left(Y - \overline{Y}\right).$$

With one addition, one subtraction, and one substitution, we can transform this equation into the Phillips curve relationship between inflation and unemployment.

Here are the three steps. First, add to the right-hand side of the equation a supply shock *v* to represent exogenous events, like a change in world oil prices, that alter the price level and shift the short-run aggregate supply curve:

$$P = EP + (1/\alpha)\left(Y - \overline{Y}\right) + v.$$

Next, to go from the price level to inflation rates, subtract last year's price level P_{-1} from both sides of the equation to obtain

$$(P - P_{-1}) = (EP - P_{-1}) + (1/\alpha)\left(Y - \overline{Y}\right) + v.$$

The term on the left-hand side, $P - P_{-1}$, is the difference between the current price level and last year's price level, which is inflation π.[6] The term on the right-hand side, $EP - P_{-1}$, is the difference between the expected price level and last year's price level, which is expected inflation $E\pi$. Therefore, we can replace $P - P_{-1}$ with π and $EP - P_{-1}$ with $E\pi$:

$$\pi = E\pi + (1/\alpha)\left(Y - \overline{Y}\right) + v.$$

Third, to go from output to unemployment, recall from Chapter 10 that Okun's law gives a relationship between these two variables. One version of Okun's law states that the deviation of output from its natural level is inversely related to the deviation of unemployment from its natural rate; that is, when output is higher than the natural level of output, unemployment is lower than the natural rate of unemployment. We can write this as

$$(1/\alpha)\left(Y-\overline{Y}\right) = -\beta\left(u-u^n\right).$$

Using this Okun's law relationship, we can substitute $-\beta\left(u-u^n\right)$ for $(1/\alpha)\left(Y-\overline{Y}\right)$ in the previous equation to obtain:

$$\pi = E\pi - \beta\left(u-u^n\right) + v.$$

Thus, we can derive the Phillips curve equation from the aggregate supply equation.

All this algebra is meant to show one thing: the Phillips curve equation and the short-run aggregate supply equation represent the same economic ideas. Both equations show a link between real and nominal variables that causes the classical dichotomy (the theoretical separation of real and nominal variables) to break down in the short run. According to the short-run aggregate supply equation, output is related to unexpected movements in the price level. According to the Phillips curve equation, unemployment is related to unexpected movements in the inflation rate. The aggregate supply curve is more convenient when studying output and the price level, whereas the Phillips curve is more convenient when studying unemployment and inflation. But always remember that the Phillips curve and the aggregate supply curve are two sides of the same coin.

FYI

The History of the Modern Phillips Curve

The Phillips curve is named after economist A. W. Phillips. In 1958 Phillips observed a negative relationship between the unemployment rate and the rate of wage inflation in data for the United Kingdom.[7] The Phillips curve that economists use today differs in three ways from the relationship Phillips examined.

First, the modern Phillips curve substitutes price inflation for wage inflation. This difference is not crucial because price inflation and wage inflation are closely related. In periods when wages are rising quickly, prices are also rising quickly.

Second, the modern Phillips curve includes expected inflation. This addition is due to the work of Milton Friedman and Edmund Phelps. In developing early versions of the imperfect-information model in the 1960s, these economists emphasized the importance of expectations for aggregate supply.

Third, the modern Phillips curve includes supply shocks. Credit for this addition goes to OPEC, the Organization of the Petroleum Exporting Countries. In the 1970s OPEC caused large increases in the world price of oil, which made economists more aware of the importance of shocks to aggregate supply.

Adaptive Expectations and Inflation Inertia

To make the Phillips curve useful for analyzing the choices facing policymakers, we need to specify what determines expected inflation. A simple and often plausible assumption is that people form their expectations of inflation based on recently observed inflation. This assumption is called **adaptive expectations**. For example, suppose people expect prices to rise this year at the same rate as they did last year. Then expected inflation $E\pi$ equals last year's inflation π_{-1}:

$$E\pi = \pi_{-1}.$$

In this case, we can write the Phillips curve as

$$\pi = \pi_{-1} - \beta(u - u^n) + v,$$

which states that inflation depends on past inflation, cyclical unemployment, and a supply shock. When the Phillips curve is written in this form, the natural rate of unemployment is sometimes called the nonaccelerating inflation rate of unemployment, or *NAIRU*.

The first term in this form of the Phillips curve, π_{-1}, implies that inflation has inertia. That is, like an object moving through space, inflation keeps going unless something acts to stop it. In particular, if unemployment is at the NAIRU and if there are no supply shocks, the increase in the price level neither speeds up nor slows down. This inertia arises because past inflation influences expected future inflation, which in turn influences the wages and prices that people set. Writing during the high inflation of the 1970s, Robert Solow offered a succinct summary of inflation inertia: "Why is our money ever less valuable? Perhaps it is simply that we have inflation because we expect inflation, and we expect inflation because we've had it."

In the model of aggregate supply and aggregate demand, inflation inertia is interpreted as persistent upward shifts in both the aggregate supply and aggregate demand curves. First, consider aggregate supply. If prices have been rising quickly, people will expect them to continue to rise quickly. Because the position of the short-run aggregate supply curve depends on the expected price level, the short-run aggregate supply curve will shift upward over time. It will continue to shift upward until some event, such as a recession or a supply shock, changes inflation and thereby changes expectations of inflation.

The aggregate demand curve must also shift upward to confirm the expectations of inflation. Most often, the continued rise in aggregate demand is due to persistent growth in the money supply. If the Fed suddenly halted money growth, aggregate demand would stabilize, and the upward shift in aggregate supply would cause a recession. The high unemployment in the recession would reduce inflation and expected inflation, causing inflation inertia to subside.

Two Causes of Rising and Falling Inflation

The second and third terms in the Phillips curve equation show the two forces that can change the rate of inflation.

The second term, $\beta(u - u^n)$, shows that cyclical unemployment—the deviation of unemployment from its natural rate—exerts upward or downward pressure on inflation. Low unemployment pulls the inflation rate up. This is called **demand-pull inflation** because high aggregate demand is responsible for this type of inflation. Conversely, high unemployment pulls the inflation rate down. The parameter β measures how responsive inflation is to cyclical unemployment.

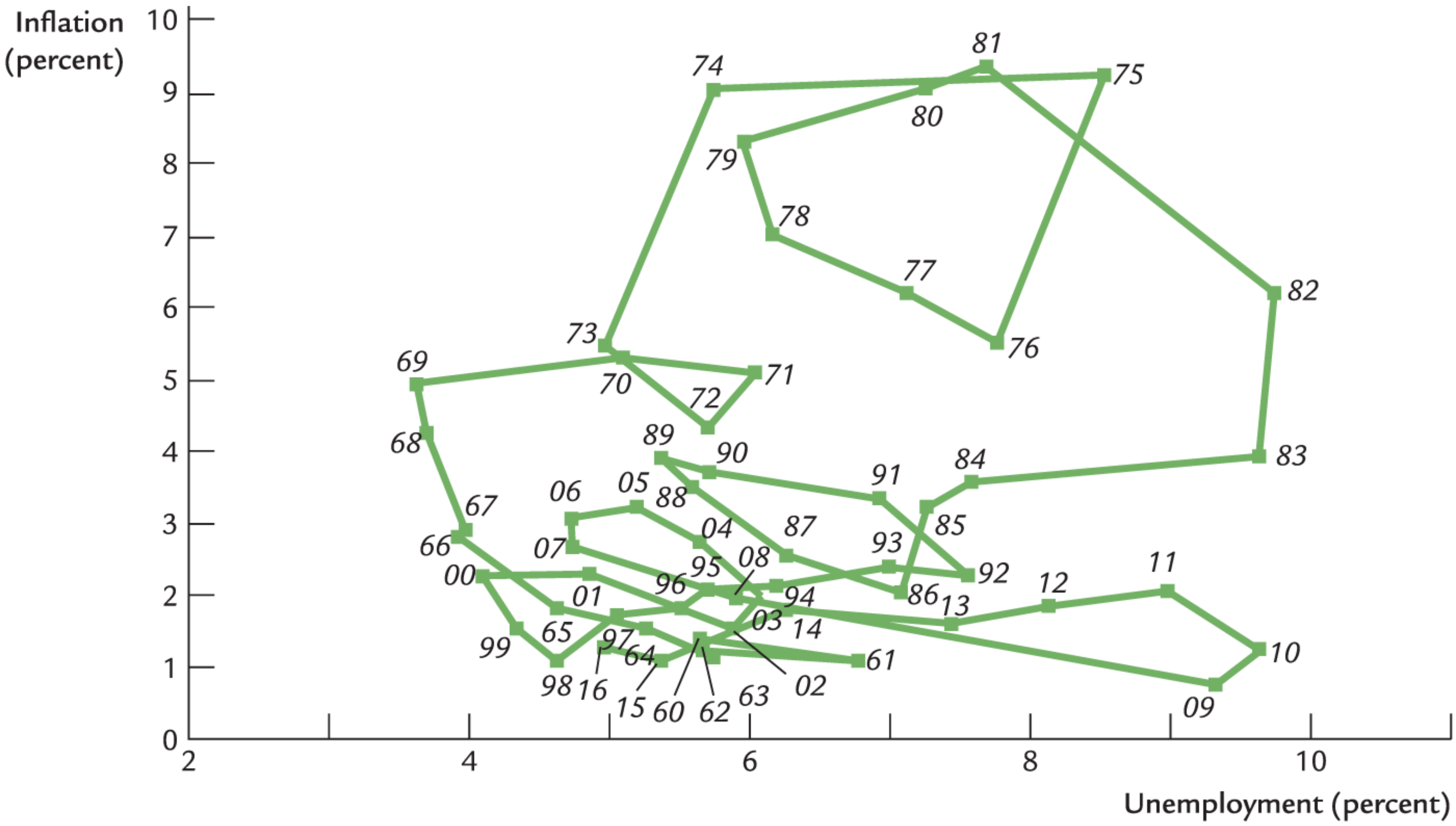
The third term, *v*, shows that inflation also rises and falls because of supply shocks. An adverse supply shock implies a positive value of *v* and causes inflation to rise. This is called **cost-push inflation** because adverse supply shocks are events that push up the costs of production. A beneficial supply shock reduces the costs of production, makes *v* negative, and causes inflation to fall.

History is full of examples of demand-pull and cost-push inflation, as the next case study shows.

CASE STUDY

Inflation and Unemployment in the United States

Because inflation and unemployment are important measures of economic performance, macroeconomic developments are often viewed through the lens of the Phillips curve. Figure 14-3 displays the history of inflation and unemployment in the United States from 1960 to 2016. These data, spanning more than half a century, illustrate some of the causes of rising or falling inflation.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 14-3 Inflation and Unemployment in the United States, 1960–2016 This figure uses annual data on the unemployment rate and the inflation rate (percentage change in the GDP deflator) to illustrate macroeconomic developments spanning half a century of U.S. history.

Data from: U.S. Department of Commerce and U.S. Department of Labor.

The 1960s showed how policymakers can, in the short run, lower unemployment, leading to demand-pull inflation. The tax cut of 1964, together with expansionary monetary policy, expanded aggregate demand and pushed the unemployment rate below 5 percent. This expansion of aggregate demand continued in the late 1960s as a byproduct of government spending for the Vietnam War. Unemployment fell lower and inflation rose higher than policymakers intended.

The 1970s were a period of economic turmoil. The decade began with policymakers trying to reduce the inflation inherited from the 1960s. President Nixon imposed temporary controls on wages and prices, and the Fed engineered a recession through contractionary monetary policy, but the inflation rate fell only slightly. The effects of wage and price controls ended when the controls were lifted, and the recession was too small to counteract the inflationary impact of the boom that had preceded it. By 1972 the unemployment rate was the same as it had been a decade earlier, while inflation was about 3 percentage points higher.

Beginning in 1973 policymakers had to cope with large supply shocks and cost-push inflation. OPEC first raised oil prices in the mid-1970s, increasing the inflation rate to above 9 percent. This adverse supply shock, together with temporarily tight monetary policy, caused a recession in 1975. High unemployment during the recession reduced inflation somewhat, but further OPEC price hikes pushed inflation up again in the late 1970s.

The 1980s began with high inflation and high expectations of inflation. Under the leadership of Chair Paul Volcker, the Fed doggedly pursued monetary policies aimed at reducing inflation. In 1982 and 1983 the unemployment rate reached its highest level in 40 years. High unemployment, aided by a fall in oil prices in 1986, pulled the inflation rate down from about 9 percent to about 2 percent. By 1987 the unemployment rate of about 6 percent was close to most estimates of the natural rate. Unemployment continued to fall through the 1980s, however, reaching a low of 5.3 percent in 1989 and beginning a new round of demand-pull inflation.

Compared to the preceding 30 years, the 1990s and early 2000s were relatively quiet. The 1990s began with a recession caused by several contractionary shocks to aggregate demand: tight monetary policy, the savings-and-loan crisis, and a fall in consumer confidence coinciding with the Gulf War. The unemployment rate rose to 7.5 percent in 1992, and inflation fell slightly. Unlike in the 1982 recession, unemployment in the 1990 recession was never far above the natural rate, so the effect on inflation was small. Similarly, a recession in 2001 (discussed in Chapter 12) raised unemployment, but the downturn was mild by historical standards, and the impact on inflation was once again slight.

A more severe recession began in 2008. As we discussed in Chapter 12, the cause of this downturn was a financial crisis, leading to a substantial decline in aggregate demand. Unemployment rose significantly in 2009, and the inflation rate fell to low levels, much as the conventional Phillips curve predicts. With unemployment so persistently high, some economists worried that the economy would experience deflation (a negative inflation rate). Yet that did not occur. One possible explanation is that expectations of inflation remained anchored at around 2 percent instead of changing as the assumption of adaptive expectations would indicate. That is, the Fed's recent history had given the central bank enough credibility about its target rate of inflation that expected inflation did not change as quickly as it might have in past episodes.

Thus, U.S. macroeconomic history illustrates the many forces working on the inflation rate, as described in the Phillips curve equation. The 1960s and 1980s show the two sides of demand-pull inflation: in the 1960s low unemployment pulled inflation up, and in the 1980s high unemployment pulled inflation down. The oil-price hikes of the 1970s show the effects of cost-push inflation. And the aftermath of the recession of 2008–2009 shows that inflation sometimes surprises us, in part because changing expectations are not always easy to predict.[8] ■

The Short-Run Tradeoff Between Inflation and Unemployment

Consider the options the Phillips curve gives to a policymaker who can influence aggregate demand with monetary or fiscal policy. At any moment, expected inflation and supply shocks are beyond his immediate control. Yet, by changing aggregate demand, the policymaker can alter output, unemployment, and inflation. He can expand aggregate demand to lower unemployment and raise inflation. Or he can depress aggregate demand to raise unemployment and lower inflation.

Figure 14-4 plots the Phillips curve equation and shows the short-run tradeoff between inflation and unemployment. When unemployment is at its natural rate (u=un),$(u = u^n)$, inflation depends on expected inflation and the supply shock (π=Eπ+v).$(\pi = E\pi + v)$. The parameter ββ determines the slope of the tradeoff between inflation and unemployment. In the short run, for a given rate of expected inflation, policymakers can manipulate aggregate demand to choose any combination of inflation and unemployment on this curve, called the *short-run Phillips curve*.

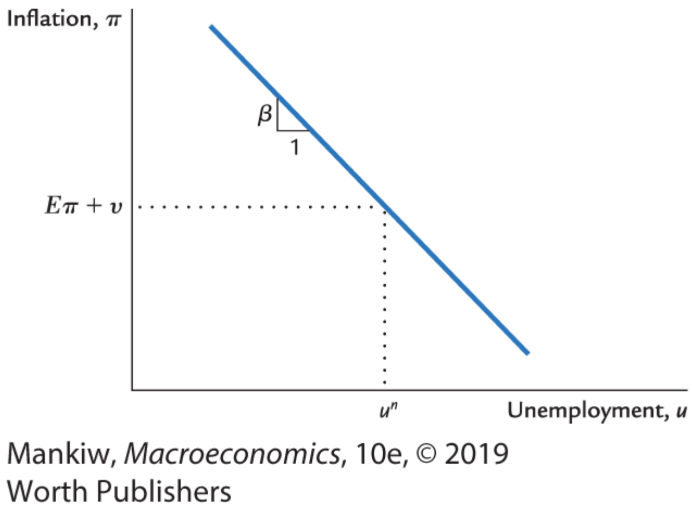


FIGURE 14-4 The Short-Run Tradeoff Between Inflation and Unemployment In the short run, inflation and unemployment are negatively related. At any point in time, a policymaker who controls aggregate demand can choose a combination of inflation and unemployment on this short-run Phillips curve.

Notice that the position of the short-run Phillips curve depends on the expected rate of inflation. If expected inflation rises, the curve shifts upward, and the policymaker's tradeoff becomes less favorable: inflation is higher for any level of unemployment. Figure 14-5 shows how the tradeoff depends on expected inflation.

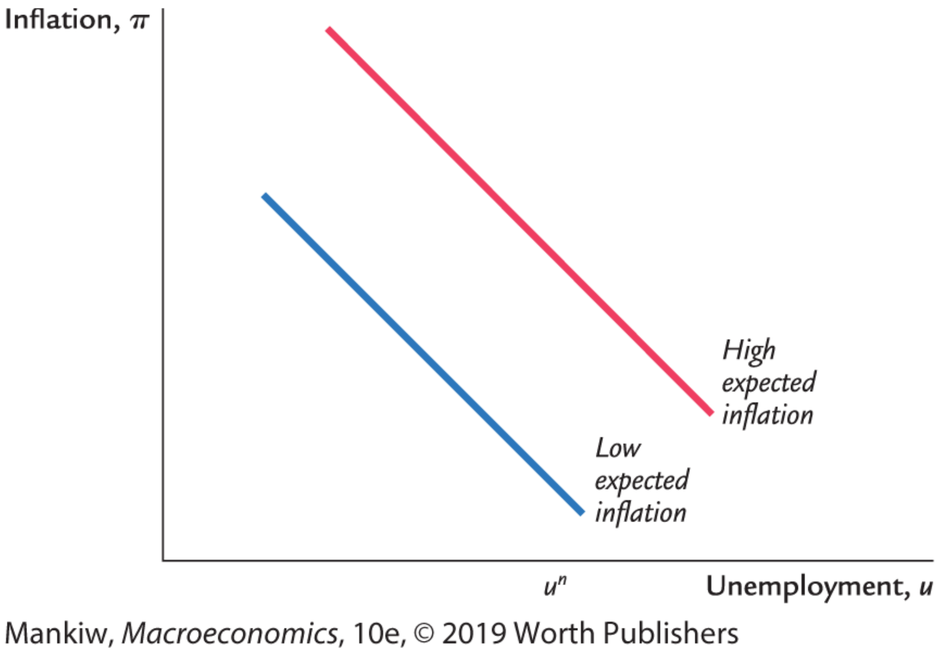


FIGURE 14-5 Shifts in the Short-Run Tradeoff The short-run tradeoff between inflation and unemployment depends on expected inflation. The curve is higher when expected inflation is higher.

Because people adjust their expectations of inflation over time, the tradeoff between inflation and unemployment holds only in the short run. The policymaker cannot keep inflation above expected inflation (and thus unemployment below its natural rate) forever. Eventually, expectations adapt to whatever inflation rate the policymaker chooses. In the long run, the classical dichotomy holds, unemployment returns to its natural rate, and there is no tradeoff between inflation and unemployment.

Disinflation and the Sacrifice Ratio

Imagine an economy in which unemployment is at its natural rate and inflation is running at 6 percent. What would happen to unemployment and output if the central bank pursued a policy to reduce inflation from 6 percent to 2 percent?

The Phillips curve shows that without a beneficial supply shock, lowering inflation requires a period of high unemployment and reduced output. But by how much and for how long would unemployment need to rise above the natural rate? Before deciding whether to reduce inflation, policymakers must know how much output would be lost during the transition to lower inflation. This cost can then be compared with the benefits of lower inflation.

Much research has used the available data to examine the Phillips curve quantitatively. The results of these studies are often summarized in a number called the **sacrifice ratio**, the percentage of a year's real GDP that must be forgone to reduce inflation by 1 percentage point. Although estimates of the sacrifice ratio vary substantially, a typical estimate is about 5: that is, for every percentage point that inflation is to fall, 5 percent of one year's GDP must be sacrificed.[10]

We can also express the sacrifice ratio in terms of unemployment. Okun's law says that a change of 1 percentage point in the unemployment rate translates into a change of 2 percentage points in GDP. Therefore, reducing inflation by 1 percentage point requires about 2.5 percentage points of cyclical unemployment.

We can use the sacrifice ratio to estimate by how much and for how long unemployment must rise to reduce inflation. If reducing inflation by 1 percentage point requires a sacrifice of 5 percent of a year's GDP, reducing inflation by 4 percentage points requires a sacrifice of 20 percent of a year's GDP. Equivalently, this reduction in inflation requires a sacrifice of 10 percentage points of cyclical unemployment.

This disinflation could take various forms, each totaling the same sacrifice of 20 percent of a year's GDP. For example, a rapid disinflation would lower output by 10 percent for two years: this is sometimes called the *cold-turkey* solution to inflation. A moderate disinflation would lower output by 5 percent for four years. An even more gradual disinflation would depress output by 2 percent for a decade.

FYI

How Precise Are Estimates of the Natural Rate of Unemployment?

If you ask an astronomer how far a particular star is from our sun, he'll give you a number, but it won't be accurate. Our ability to measure astronomical distances is limited. An astronomer might take better measurements and conclude that a star is really twice or half as far away as he previously thought.

Estimates of the natural rate of unemployment, or NAIRU, are also far from precise. One problem is supply shocks. Shocks to oil supplies, crop harvests, or technological progress can cause inflation to rise or fall in the short run. When we observe rising inflation, therefore, we cannot be sure whether it is evidence that the unemployment rate is below the natural rate or evidence that the economy is experiencing an adverse supply shock.

A second problem is that the natural rate is not constant over time. Demographic changes (such as the aging of the baby-boom generation), policy changes (such as minimum-wage laws), and institutional changes (such as the declining role of unions) all affect the normal level of unemployment. Estimating the natural rate is like trying to hit a moving target.

Economists deal with these problems using statistical techniques that yield a best guess about the natural rate and allow them to gauge the uncertainty associated with their estimates. In one study, Douglas Staiger, James Stock, and Mark Watson estimated the natural rate to be 6.2 percent in 1990, with a 95 percent confidence interval between 5.1 and 7.7 percent. A 95 percent confidence interval is a range such that the statistician is 95 percent confident that the true value falls in that range. A more recent study by economists at the Federal Reserve put the natural rate in 2013 at 5.8 percent, with a 95 percent confidence interval between 4.5 and 7 percent. These large confidence intervals show that estimates of the natural rate of unemployment are not at all precise.

This conclusion has profound implications. Policymakers may want to keep unemployment close to its natural rate, but their ability to do so is limited by the fact that they cannot be sure what that natural rate is.[9]

Rational Expectations and the Possibility of Painless Disinflation

Because expected inflation influences the short-run tradeoff between inflation and unemployment, it is crucial to understand how people form expectations. So far, we have assumed that expected inflation depends on recently observed inflation. This assumption of adaptive expectations is plausible, but it may be too simple to apply in all circumstances.

An alternative approach is to assume that people have **rational expectations**. That is, we might assume that people optimally use all available information, including information about government policies, to forecast the future. Because monetary and fiscal policies influence inflation, expected inflation should also depend on the monetary and fiscal policies in effect. According to the theory of rational expectations, a change in monetary or fiscal policy will change expectations, and an evaluation of any policy change must incorporate this effect on expectations. If people form their expectations rationally, inflation may be less inertial than it appears.

Here is how Thomas Sargent, an advocate of rational expectations and a Nobel laureate in economics, describes its implications for the Phillips curve:

> An alternative “rational expectations” view denies that there is any inherent momentum in the present process of inflation. This view maintains that firms and workers have now come to expect high rates of inflation in the future and that they strike inflationary bargains in light of these expectations. However, it is held that people expect high rates of inflation in the future precisely because the government’s current and prospective monetary and fiscal policies warrant those expectations. . . . Thus inflation only seems to have a momentum of its own; it is actually the long-term government policy of persistently running large deficits and creating money at high rates which imparts the momentum to the inflation rate. An implication of this view is that inflation can be stopped much more quickly than advocates of the “momentum” view have indicated and that their estimates of the length of time and the costs of stopping inflation in terms of foregone output are erroneous. . . . [Stopping inflation] would require a change in the policy regime: there must be an abrupt change in the continuing government policy, or strategy, for setting deficits now and in the future that is sufficiently binding as to be widely believed. . . . How costly such a move would be in terms of foregone output and how long it would be in taking effect would depend partly on how resolute and evident the government’s commitment was.[11]

Thus, advocates of rational expectations argue that the short-run Phillips curve does not accurately represent the options that policymakers have available. They believe that if policymakers are credibly committed to reducing inflation, rational people will understand the commitment and will quickly lower their expectations of inflation. Inflation can then decline without a rise in unemployment and fall in output. According to the theory of rational expectations, traditional estimates of the sacrifice ratio are not useful for evaluating the impacts of alternative policies. Under a credible policy, the costs of reducing inflation may be much lower than estimates of the sacrifice ratio suggest.

In the most extreme case, policymakers can reduce inflation without causing any recession at all. A painless disinflation has two requirements. First, the plan to reduce inflation must be announced before the workers and firms that set wages and prices have formed their expectations. Second, the workers and firms must believe the announcement; otherwise, their expectations of inflation will not fall. If both requirements are met, the announcement will quickly shift the short-run tradeoff between inflation and unemployment downward, permitting lower inflation without higher unemployment.

The rational-expectations approach remains controversial, but most economists agree that expectations of inflation influence the short-run tradeoff between inflation and unemployment. The credibility of a policy to reduce inflation is therefore one determinant of how costly the policy will be. Because it is hard to know whether the public will view the announcement of a new policy as credible, the central role of expectations makes predicting the results of alternative policies more difficult.

CASE STUDY

The Sacrifice Ratio in Practice

The Phillips curve with adaptive expectations implies that reducing inflation requires a period of high unemployment and low output. By contrast, the rational-expectations approach suggests that reducing inflation can be much less costly. What happens during actual disinflations?

Consider the U.S. disinflation in the early 1980s. This decade began with some of the highest rates of inflation in U.S. history. Yet because of the tight monetary policies the Fed pursued under Chair Paul Volcker, the rate of inflation fell substantially in the first few years of the decade. This episode provides a natural experiment with which to estimate how much output is lost during the process of disinflation.

The first question is, how much did inflation fall? As measured by the GDP deflator, inflation reached a peak of 9.3 percent in 1981. It is natural to end the episode in 1985 because oil prices plunged in 1986—a large, beneficial supply shock unrelated to Fed policy. In 1985, inflation was 3.2 percent, so we can estimate that the Fed engineered a reduction in inflation of 6.1 percentage points over four years.

The second question is, how much output was lost during this period? Table 14-1 shows the unemployment rate from 1982 to 1985. Assuming that the natural rate of unemployment was 6 percent, we can compute the amount of cyclical unemployment in each year. In total over this period, there were 10.0 percentage points of cyclical unemployment. Okun's law says that 1 percentage point of unemployment translates into 2 percentage points of GDP. Therefore, 20.0 percentage points of annual GDP were lost during the disinflation.

TABLE 14-1 Unemployment During the Volcker Disinflation

Year	Unemployment Rate u	Natural Rate u^n	Cyclical Unemployment $u - u^n$
1982	9.7%	6.0%	3.7%
1983	9.6	6.0	3.6
1984	7.5	6.0	1.5
1985	7.2	6.0	1.2
			Total 10.0%

Now we can compute the sacrifice ratio for this episode. We know that 20.0 percentage points of GDP were lost and that inflation fell by 6.1 percentage points. Hence, 20.0/6.1, or 3.3, percentage points of GDP were lost for each percentage-point reduction in inflation. The estimate of the sacrifice ratio from the Volcker disinflation is 3.3.

This estimate of the sacrifice ratio is smaller than the estimates made before Volcker was appointed Fed chair. In other words, Volcker reduced inflation at a smaller cost than many economists had predicted. One explanation is that Volcker's tough stand was credible enough to influence expectations of inflation directly. Yet the change in expectations was not large enough to make the disinflation painless: in 1982 unemployment reached 10.8 percent, its highest level since the Great Depression.

The Volcker disinflation is only one historical episode, but this kind of analysis can be applied to other disinflations. One comprehensive study documented the results of 65 disinflations in 19 countries. In almost all cases, the reduction in inflation came at the cost of temporarily lower output. Yet the size of the output loss varied from episode to episode. Rapid disinflations usually had smaller sacrifice ratios than slower ones. That is, in contrast to what the Phillips curve with adaptive expectations suggests, a cold-turkey approach appears less costly than a gradual one. Moreover, countries with more flexible wage-setting institutions, such as shorter labor contracts, had smaller sacrifice ratios. These findings indicate that reducing inflation always has some cost but that policies and institutions can affect its magnitude.[12] ■

Hysteresis and the Challenge to the Natural-Rate Hypothesis

Our discussion of the cost of disinflation—and indeed our entire discussion of economic fluctuations in the past four chapters—has been based on an assumption called the **natural-rate hypothesis**. This hypothesis is summarized in the following statement:

> *Fluctuations in aggregate demand affect output and employment only in the short run. In the long run, the economy returns to the levels of output, employment, and unemployment described by the classical model.*

The natural-rate hypothesis allows macroeconomists to separately study short-run and long-run developments in the economy. It is one expression of the classical dichotomy.

Some economists, however, have challenged the natural-rate hypothesis by suggesting that aggregate demand may affect output and employment even in the long run. They have pointed out a number of mechanisms through which recessions might leave permanent scars on the economy by altering the natural rate of unemployment. **Hysteresis** is the term used to describe the long-lasting influence of history on the natural rate.

A recession can have permanent effects if it changes the people who become unemployed. For instance, workers might lose valuable job skills when unemployed, diminishing their ability to find a job even after the recession ends. Or a long period of unemployment may change a person's attitude toward work and reduce his desire to find employment. In either case, the recession permanently inhibits the process of job search and increases frictional unemployment.

Another way in which a recession can permanently affect the economy is by changing the process that determines wages. Those who become unemployed may lose their influence on the wage-setting process. Unemployed workers may lose their status as union members, for example. More generally, some of the *insiders* in the wage-setting process become *outsiders*. If the smaller group of insiders cares more about high real wages and less about high employment, the recession may permanently push real wages farther above the equilibrium level and increase structural unemployment.

Hysteresis remains controversial. Some economists believe this theory helps explain persistently high unemployment in Europe. The rise in European unemployment, which started in the early 1980s, coincided with disinflation but continued after inflation stabilized. Moreover, the increase in unemployment tended to be larger for the countries that experienced the greatest reductions in inflations, such as Ireland, Italy, and Spain. As these episodes suggest, hysteresis can increase the sacrifice ratio because output is lost even after the period of disinflation is over. Yet there is still no consensus on whether the hysteresis phenomenon is significant or why it might be more pronounced in some countries than in others. (Alternative explanations of high European unemployment, discussed in Chapter 7, point to forces other than disinflation.) If the theory of hysteresis is true, however, it is important because it greatly increases the cost of recessions.[13]

14-3 Conclusion

We began this chapter by discussing two models of aggregate supply, each of which focuses on a particular reason output rises above its natural level in the short run when the price level rises above the level that people had expected. Both models explain why the short-run aggregate supply curve slopes upward, and both yield a short-run tradeoff between inflation and unemployment. A convenient way to express that tradeoff is with the Phillips curve equation, according to which inflation depends on expected inflation, cyclical unemployment, and supply shocks.

Not all economists endorse every idea discussed here. There is disagreement, for instance, about the practical relevance of rational expectations and hysteresis. If you find it hard to fit all the pieces together, you are not alone. The study of aggregate supply remains one of the most unsettled—and therefore one of the most exciting—research areas in macroeconomics.

APPENDIX

The Mother of All Models

In the previous chapters, we have seen many models of how the economy works. When learning these models, it can be hard to see how they are related. Now that we have finished developing the model of aggregate demand and aggregate supply, this is a good time to review what we have learned. This appendix sketches a large model that incorporates much of the theory we have already seen, including the classical theory presented in Part Two and the business cycle theory presented in Part Four. The notation and equations should be familiar. The goal is to put much of our previous analysis into a common framework to clarify the relationships among the various models.

This comprehensive model has seven equations:

Y=C(Y−T)+I(r)+G+NX(ε)IS: Goods Market EquilibriumM/P=L(i, Y)LM: Money Market EquilibriumNX
Foreign-Exchange-
Market Equilibriumi=r+EπRelationship Between Real and Nominal Interest Ratesε=eP/P Relationship Bet
Output

$$
\begin{aligned}
Y &= C(Y-T) + I(r) + G + NX(\varepsilon) && IS: \text{ Goods Market Equilibrium} \\
M/P &= L(i,\ Y) && LM: \text{ Money Market Equilibrium} \\
NX(\varepsilon) &= CF(r - r^*) && \text{Foreign-Exchange-Market Equilibrium} \\
i &= r + E\pi && \text{Relationship Between Real and Nominal Interest Rates} \\
\varepsilon &= eP/P^* && \text{Relationship Between Real and Nominal Exchange Rates} \\
Y &= \overline{Y} + \alpha(P - EP) && \text{Aggregate Supply} \\
\overline{Y} &= F(\overline{K},\ \overline{L}) && \text{Natural Level of Output}
\end{aligned}
$$

These seven equations determine the equilibrium values of seven endogenous variables: output *Y*, the natural level of output Y$\overline{Y}$, the real interest rate *r*, the nominal interest rate *i*, the real exchange rate εε, the nominal exchange rate *e*, and the price level *P*.

Many exogenous variables influence these endogenous variables. They include the money supply *M*, government purchases *G*, taxes *T*, the capital stock *K*, the labor force *L*, the world price level *P* , and the world real interest rate *r* . In addition, there are two expectation variables: the expectation of future inflation Eπ$E\pi$ and the expectation of the current price level formed in the past *EP*. As written, the model takes these expectations as exogenous, although equations could be added to make them endogenous.

The mathematical techniques available to analyze this seven-equation model are beyond the scope of this book. But this large model is still useful because it shows how the smaller models we have examined are related to one another. In particular, *many of the models we have been studying are special cases of this large model*. Let's consider six special cases. (A problem at the end of this section asks you to examine a few more.)

Special Case 1: The Classical Closed Economy

Suppose that $EP = P,\ L(i,\ Y) = (1/V)Y,$ and $CF(r - r^*) = 0.$ In words, these equations mean that expectations of the price level adjust so that expectations are correct, money demand is proportional to income, and there are no international capital flows. In this case, output is always at its natural level, the real interest rate adjusts to equilibrate the goods market, the price level moves parallel with the money supply, and the nominal interest rate adjusts one for one with expected inflation. This special case corresponds to the economy analyzed in Chapters 3 and 5.

Special Case 2: The Classical Small Open Economy

Suppose that $EP = P,\ L(i,\ Y) = (1/V)Y,$ and $CF(r - r^*)$ is infinitely elastic. Now, international capital flows respond greatly to any differences between the domestic and world interest rates. This means that $r = r^*$ and that the trade balance *NX* equals the difference between saving and investment at the world interest rate. This special case corresponds to the economy analyzed in Chapter 6.

Special Case 3: The Basic Model of Aggregate Demand and Aggregate Supply

Suppose that α is infinite and $L(i,\ Y) = (1/V)Y.$ In this case, the short-run aggregate supply curve is horizontal, and the aggregate demand curve is determined only by the quantity equation. This special case corresponds to the economy analyzed in Chapter 10.

Special Case 4: The *IS–LM* Model

Suppose that α is infinite and $CF(r - r^*) = 0.$ Now, the short-run aggregate supply curve is horizontal, and there are no international capital flows. For any given rate of expected inflation $E\pi,$ income and the interest rate must adjust to equilibrate the goods market and the money market. This special case corresponds to the economy analyzed in Chapters 11 and 12.

Special Case 5: The Mundell–Fleming Model with a Floating Exchange Rate

Suppose that α is infinite and $CF(r - r^*)$ is infinitely elastic. In this case, the short-run aggregate supply curve is horizontal, and international capital flows are so great as to ensure that $r = r^*$.

$r = r^*$. The exchange rate floats freely to reach its equilibrium level. This special case corresponds to the first economy analyzed in Chapter 13.

Special Case 6: The Mundell–Fleming Model with a Fixed Exchange Rate

Suppose that α is infinite, $CF(r - r^*)$ is infinitely elastic, and the nominal exchange rate *e* is fixed. In this case, the short-run aggregate supply curve is horizontal, huge international capital flows ensure that $r = r^*$, but the exchange rate is set by the central bank. The exchange rate is now an exogenous policy variable, but the money supply *M* is an endogenous variable that must adjust to ensure the exchange rate hits the fixed level. This special case corresponds to the second economy analyzed in Chapter 13.

You should now see the value in this big model. Even though the model is too large to be useful in developing an intuitive understanding of how the economy works, it shows that the different models we have been studying are closely related. In each chapter, we made simplifying assumptions to make the big model smaller and easier to understand.

Figure 14-6 presents a schematic diagram that depicts how the various models are related. It shows how, starting with the mother of all models above, you can arrive at some of the models examined in previous chapters. Here are the steps:

1. *Classical or Keynesian?* You decide whether you want a classical special case (which occurs when $EP = P$ or when α equals zero, so output is at its natural level) or a Keynesian special case (which occurs when α equals infinity, so the price level is completely fixed).
2. *Closed or Open?* You decide whether you want a closed economy (which occurs when the capital flow *CF* always equals zero) or an open economy (which allows *CF* to differ from zero).
3. *Small or Large?* If you want an open economy, you decide whether you want a small one (in which *CF* is infinitely elastic at the world interest rate r^*) or a large one (in which the domestic interest rate is not pinned down by the world rate).
4. *Floating or Fixed?* If you are examining a small open economy, you decide whether the exchange rate is floating (in which case the central bank sets the money supply) or fixed (in which case the central bank allows the money supply to adjust).
5. *Fixed Velocity?* If you are considering a closed economy with the Keynesian assumption of fixed prices, you decide whether you want to focus on the special case in which velocity is exogenously fixed.

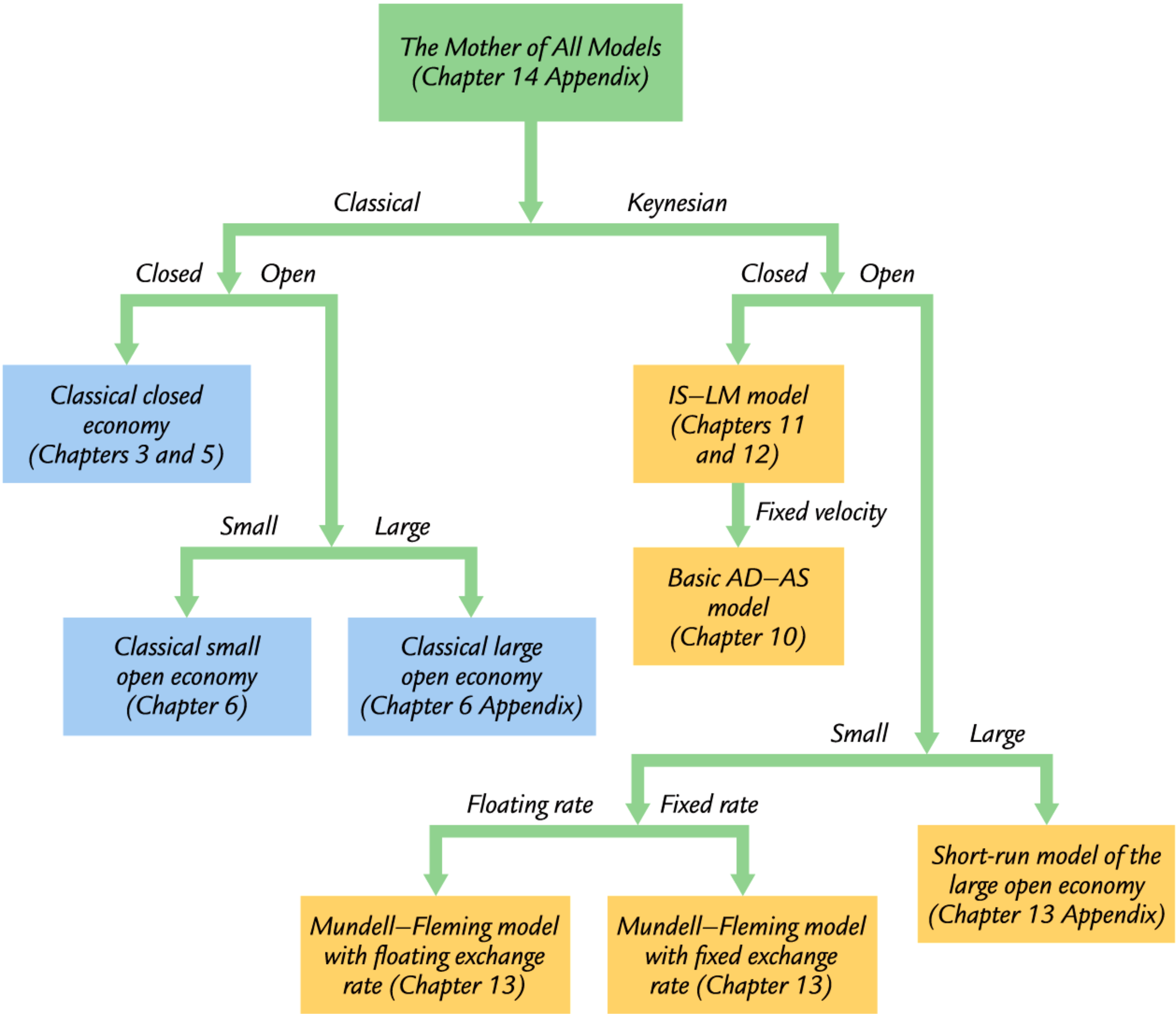


FIGURE 14-6 How Models Are Related This schematic diagram shows how the large, comprehensive model presented in this appendix is related to the smaller, simpler models developed in earlier chapters.

By making this series of modeling decisions, you move from the more complete and complex model to a simpler, more narrowly focused special case that is easier to understand and use.

When thinking about the real world, you should keep in mind all the models and their simplifying assumptions. Each model provides insight into some facet of the economy.

MORE PROBLEMS AND APPLICATIONS

1. Let's consider some more special cases of the mother of all models. Starting with this comprehensive model, what extra assumptions would you need to yield each of the following specialized models?
    a. The model of the classical large open economy in the appendix to Chapter 6
    b. The Keynesian cross in the first half of Chapter 11
    c. The *IS–LM* model for the large open economy in the appendix to Chapter 13

CHAPTER 15

A Dynamic Model of Economic Fluctuations

> The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.
>
> —William Bragg

The opening quotation from William Bragg (a physicist who lived about a century ago) applies just as much to economics as it does to the natural sciences. Many of the facts that economists study are reported by the media every day—GDP, inflation, unemployment, the trade balance, and so on. Economists develop models to provide new ways to think about these familiar facts. A good model is one that not only fits the facts but also offers new insights.

In the previous chapters, we developed models that explain the economy in both the long run and the short run. It might seem that, in some sense, our study of macroeconomics is complete. But like all other scientists, economists never rest. There are always more questions to be answered and more refinements to be made. In this chapter and the next four, we look at some topics in macroeconomic theory and policy that expand our understanding of the economy and the choices facing policymakers.

This chapter presents a model that we will call the *dynamic model of aggregate demand and aggregate supply*. This model offers another lens through which we can view short-run fluctuations in output and inflation and the effects of monetary and fiscal policy on those fluctuations. As the name suggests, this new model emphasizes the dynamic nature of economic fluctuations. The dictionary defines the word *dynamic* as "relating to energy or objects in motion; characterized by continuous change or activity." This definition applies readily to economic activity. The economy is continually bombarded by various shocks. These shocks not only have an immediate impact on the economy's short-run equilibrium but also affect the subsequent path of output, inflation, and many other variables. The dynamic *AD–AS* model focuses attention on how output and inflation respond over time to changes in the economic environment.

In addition to placing greater emphasis on dynamics, the model differs from our previous models in another significant way: it explicitly incorporates the response of monetary policy to economic conditions. In previous chapters, we followed the conventional simplification that the central bank sets the money supply, which in turn is one determinant of the equilibrium interest rate. In the real world, however, many central banks set a target for the interest rate and allow the money supply to adjust to the level necessary to achieve that target. Moreover, the target interest rate set by the central bank depends on economic conditions, including both inflation and output. The dynamic *AD–AS* model includes these realistic features of monetary policy.

Many of the building blocks of the dynamic *AD–AS* model will be familiar from previous chapters, though they sometimes take on slightly different forms. More importantly, these components are assembled in new ways. You can think of this model as a new recipe that mixes familiar ingredients to create a surprisingly original meal. In this case, we will mix familiar economic relationships in a new way to produce deeper insights into the nature of short-run economic fluctuations.

Compared to the models in preceding chapters, the dynamic *AD–AS* model is closer to those studied by economists at the research frontier. Moreover, economists involved in setting macroeconomic policy, including those working in central banks around the world, often use versions of this model when analyzing the impact of economic events on output and inflation.

15-1 Elements of the Model

Before examining the components of the dynamic *AD–AS* model, we need to introduce one piece of notation: throughout this chapter, the subscript *t* on a variable represents time. For example, *Y* continues to represent total output and national income, but it now takes the form Y_t, representing output in time period *t*. Similarly, Y_{t-1} represents output in period $t-1$, and Y_{t+1} represents output in period $t+1$. This new notation allows us to keep track of variables as they change over time.

Let's now look at the five equations that make up the dynamic *AD–AS* model.

Output: The Demand for Goods and Services

The demand for goods and services is given by the equation

$$Y_t = \overline{Y}_t - \alpha(r_t - \rho) + \varepsilon_t,$$

where Y_t is the total output of goods and services, $\overline{Y}_t$ is the economy's natural level of output, r_t is the real interest rate, ε_t is a random demand shock, and α and ρ are parameters greater than zero (to be explained shortly). This equation is similar in spirit to the demand for goods and services equation in Chapter 3 and the *IS* equation in Chapter 11. Because this equation is so central to the dynamic *AD–AS* model, it is important to carefully examine each term.

The first term on the right-hand side of the equation, $\overline{Y}_t$, implies that the demand for goods and services Y_t rises with the economy's natural level of output $\overline{Y}_t$. In most cases, we can simplify the analysis by assuming that $\overline{Y}_t$ is constant (that is, the same for every time period *t*). Later in the chapter, however, we examine how this model can take into account long-run growth, represented by exogenous increases in $\overline{Y}_t$ over time. Holding other things constant, as long-run growth increases the economy's ability to supply goods and services (measured by the natural level of output $\overline{Y}_t$), it also makes the economy richer and increases the demand for goods and services.

The second term on the right-hand side of the equation expresses a negative relationship between the real interest rate r_t and the demand for goods and services Y_t. When the real interest rate increases, borrowing becomes more expensive, and saving yields a greater reward. As a result, firms engage in fewer investment projects, and consumers save more and spend less. Both of these effects reduce the demand for goods and services. The parameter α tells us how sensitive demand is to changes in the real interest rate. The larger the value of α, the more the demand for goods and services responds to a given change in the real interest rate. Note that the interest rate enters this equation as $r_t - \rho$, the deviation from the parameter ρ, which we will interpret in a moment.

The last term in the demand equation, ε_t, represents exogenous shifts in demand. Think of ε_t as a *random variable*—a variable whose values are determined by chance. It is zero on average but fluctuates over time. For example, if (as Keynes famously suggested) investors are driven in part by “animal spirits”—irrational waves of optimism and pessimism—those changes in sentiment would be captured by ε_t. When investors become optimistic, they increase their demand for goods and services, represented here by a positive value of ε_t. When they become pessimistic, they cut back on spending, and ε_t is negative.

Now consider the parameter ρ. We call ρ the *natural rate of interest* because it is the real interest rate at which, in the absence of any shock, the demand for goods and services equals the natural level of output. That is, if $\varepsilon_t = 0$ and $r_t = \rho$, then $Y_t = \overline{Y}_t$. Later in the chapter, we see that the real interest rate r_t tends to move toward the natural rate of interest ρ in the long run. Throughout this chapter, we assume that the natural rate of interest is constant (that is, the same in every period). Problem 7 at the end of the chapter examines what happens if it changes.

Finally, a word about how monetary and fiscal policies influence the demand for goods and services. Monetary policymakers affect demand by changing the real interest rate r_t. Thus, their actions work through the second term in this equation. By contrast, when fiscal policymakers alter taxes or government spending, they change demand at any given interest rate. As a result, the variable ε_t captures changes in fiscal policy. An increase in government spending or a tax cut that stimulates consumer spending means a positive value of ε_t. A cut in government spending or a tax hike means a negative value of ε_t. As we will see, one purpose of this model is to examine the dynamic effects of changes in monetary and fiscal policy.

The Real Interest Rate: The Fisher Equation

The real interest rate in this model is defined as it has been in earlier chapters. The real interest rate r_t is the nominal interest rate i_t minus the expected rate of future inflation $E_t\pi_{t+1}$. That is,

$$r_t = i_t - E_t\pi_{t+1}.$$

This Fisher equation is like the one we saw in Chapter 5. Here, $E_t\pi_{t+1}$ represents the expectation

$E_t\pi_{t+1}$

formed in period *t* of inflation in period $t+1$. The variable r_t is the ex ante real interest rate: the real interest rate that people anticipate based on their expectation of inflation.

A word on the notation and timing convention should clarify the meaning of these variables. The variables r_t and i_t are interest rates that prevail at time *t* and, therefore, represent a rate of return between periods *t* and $t+1$. The variable π_t denotes the current inflation rate, which is the percentage change in the price level between periods $t-1$ and *t*. Similarly, π_{t+1} is the percentage change in the price level that will occur between periods *t* and $t+1$. As of period *t*, π_{t+1} represents a future inflation rate and therefore is not yet known. In period *t*, people can form an expectation of π_{t+1} (written as $E_t\pi_{t+1}$), but they will have to wait until period $t+1$ to learn the actual value of π_{t+1} and whether their expectation was correct.

Note that the subscript on a variable tells us when the variable is determined. The nominal and ex ante real interest rates between *t* and $t+1$ are known at time *t*, so they are written as i_t and r_t. By contrast, the inflation rate between *t* and $t+1$ is not known until time $t+1$, so it is written as π_{t+1}.

This subscript rule also applies when the expectations operator *E* precedes a variable, but here you have to be especially careful. As in previous chapters, the operator *E* in front of a variable denotes the expectation of that variable prior to its realization. The subscript on the expectations operator tells us when that expectation is formed. So $E_t\pi_{t+1}$ is the expectation of what the inflation rate will be in period $t+1$ (the subscript on π) based on information available in period *t* (the subscript on *E*). While the inflation rate π_{t+1} is not known until period $t+1$, the expectation of future inflation, $E_t\pi_{t+1}$, is known at period *t*. As a result, even though the ex post real interest rate, which is given by $i_t - \pi_{t+1}$, will not be known until period $t+1$, the ex ante real interest rate, $r_t = i_t - E_t\pi_{t+1}$, is known at time *t*.

Inflation: The Phillips Curve

Inflation in this economy is determined by a conventional Phillips curve augmented to include roles for expected inflation and exogenous supply shocks. The equation for inflation is

$$\pi_t = E_{t-1}\pi_t + \phi(Y_t - \overline{Y}_t) + \nu_t.$$

This piece of the model is similar to the Phillips curve and short-run aggregate supply equation introduced in

Chapter 14. According to this equation, inflation π_t depends on previously expected inflation $E_{t-1}\pi_t$, the deviation of output from its natural level $(Y_t - \bar{Y}_t)$, and an exogenous supply shock ν_t.

Inflation depends on expected inflation because some firms set prices in advance. When these firms expect high inflation, they anticipate that their costs will be rising quickly and that their competitors will be implementing large price hikes. The expectation of high inflation induces these firms to announce significant price increases for their own products. These price increases cause high actual inflation. Conversely, when firms expect low inflation, they forecast that costs and competitors' prices will rise only modestly. In this case, they keep their own price increases down, leading to low actual inflation.

The parameter ϕ, which is greater than zero, tells us how much inflation responds when output fluctuates around its natural level. Other things equal, when the economy is booming and output rises above its natural level $(Y_t > \bar{Y}_t)$, firms experience increasing marginal cost, so they raise prices; these price hikes increase inflation π_t. When the economy is in a slump and output is below its natural level $(Y_t < \bar{Y}_t)$, marginal cost falls, and firms cut prices; these price cuts reduce inflation π_t. The parameter ϕ reflects both how much marginal cost responds to the state of economic activity and how quickly firms adjust prices in response to changes in cost.

In this model, the state of the business cycle is measured by the deviation of output from its natural level $(Y_t - \bar{Y}_t)$. The Phillips curves in Chapter 14 sometimes emphasized the deviation of unemployment from its natural rate. This difference, however, is not significant. Recall Okun's law from Chapter 10: short-run fluctuations in output and unemployment are strongly and negatively correlated. When output is above its natural level, unemployment is below its natural rate, and vice versa. As we continue to develop this model, keep in mind that unemployment fluctuates along with output but in the opposite direction.

The supply shock ν_t is a random variable that averages to zero but can, in any given period, be positive or negative. This variable captures all influences on inflation other than expected inflation (which is captured in the first term, $E_{t-1}\pi_t$ and short-run economic conditions [which are captured in the second term, $\phi(Y_t - \bar{Y}_t)$]. For example, if an aggressive oil cartel pushes up world oil prices, thus increasing overall inflation, that event would be represented by a positive value of ν_t. If cooperation within the oil cartel breaks down and world oil prices plummet, causing inflation to fall, ν_t would be negative. In short, ν_t reflects all exogenous events that directly influence inflation.

Expected Inflation: Adaptive Expectations

As we have seen, expected inflation plays a key role in both the Phillips curve equation for inflation and the Fisher equation relating nominal and real interest rates. To keep the dynamic *AD–AS* model simple, we assume that people form their expectations of inflation based on the inflation they have recently observed. That is, people expect prices to continue rising at the same rate they have been rising. As noted in Chapter 14, this is sometimes called the assumption of *adaptive expectations*. It can be written as

$$E_t \pi_{t+1} = \pi_t .$$

When forecasting in period *t* what inflation rate will prevail in period $t + 1$, people look at inflation in period *t* and extrapolate it forward.

The same assumption applies in every period. Thus, when inflation was observed in period $t - 1$, people expected that rate to continue. This implies that $E_{t-1} \pi_t = \pi_{t-1}$.

This assumption about inflation expectations is admittedly crude. Many people are probably more sophisticated in forming their expectations. As we discussed in Chapter 14, some economists advocate an approach called *rational expectations*, according to which people optimally use all available information when forecasting the future. Incorporating rational expectations into the model is, however, beyond the scope of this book. (Moreover, the empirical validity of rational expectations is open to dispute.) The assumption of adaptive expectations simplifies the theory without losing many of its insights.

The Nominal Interest Rate: The Monetary-Policy Rule

The last piece of the model is the equation for monetary policy. We assume that the central bank sets a target for the nominal interest rate i_t based on inflation and output using this rule:

$$i_t = \pi_t + \rho + \theta_\pi (\pi_t - \pi_t^*) + \theta_Y (Y_t - \overline{Y}_t) .$$

In this equation, π_t^* is the central bank's target for the inflation rate. (For most purposes, target inflation can be assumed to be constant, but we keep a time subscript on this variable so we can later examine what happens when the central bank changes its target.) Two key policy parameters are θ_π and θ_Y, which are both assumed to be greater than zero. They indicate how much the central bank adjusts its interest rate target to changing economic conditions. The larger the value of θ_π, the more responsive the central bank is to the deviation of inflation from its target; the larger the value of θ_Y, the more responsive the central bank is to the deviation of output from its natural level. Recall that ρ, the constant in this equation, is the *natural rate of interest* (the real interest rate at which, in the absence of any shock, the demand for goods and services equals the natural level of output). This equation describes how the central bank uses monetary policy to respond to any situation it faces. In particular, it tells us how inflation and output determine the central bank's target for the nominal interest rate.

To interpret this equation, it is best to focus not only on the nominal interest rate i_t but also on the real interest rate r_t. Recall that the demand for goods and services depends on the real interest rate, not on the nominal interest rate. So, although the central bank sets a target for the nominal interest rate i_t, the bank's influence on the economy works through the real interest rate r_t. By definition, the real interest rate is $r_t = i_t - E_t\pi_{t+1}$, but with our expectation equation $E_t\pi_{t+1} = \pi_t$, we can also write the real interest rate as $r_t = i_t - \pi_t$. According to the equation for monetary policy, if inflation is at its target $(\pi_t = \pi_t^*)$ and output is at its natural level $(Y_t = \overline{Y}_t)$, the last two terms in the equation are zero, so the real interest rate equals the natural rate of interest ρ. As inflation rises above its target $(\pi_t > \pi_t^*)$ or output rises above its natural level $(Y_t > \overline{Y}_t)$, the real interest rate rises. And as inflation falls below its target $(\pi_t < \pi_t^*)$ or output falls below its natural level $(Y_t < \overline{Y}_t)$, the real interest rate falls.

At this point, one might ask, "What about the money supply?" In previous chapters, such as Chapters 11 and 12, the money supply was often taken to be the central bank's policy instrument, with the interest rate moving to bring money supply and money demand into equilibrium. Here, we turn that logic on its head. The central bank is assumed to set a target for the nominal interest rate. It then adjusts the money supply to whatever level is necessary to ensure that the equilibrium interest rate (which balances money supply and demand) hits the target.

The advantage of using the interest rate, rather than the money supply, as the policy instrument in the dynamic *AD–AS* model is that it is more realistic. Today, most central banks, including the Fed, set a short-term target for the nominal interest rate. Keep in mind, though, that hitting that target requires adjustments in the money supply. For this model, we do not need to specify the equilibrium condition for the money market, but we should remember that it is lurking in the background. When a central bank decides to change the interest rate, it is also committing itself to adjusting the money supply accordingly.

CASE STUDY

The Taylor Rule

If you wanted to set interest rates to achieve low, stable inflation while avoiding large fluctuations in output and employment, how would you do it? This is the question that Fed governors must consider every day. The short-term policy instrument that the Fed now sets is the *federal funds rate*—the short-term interest rate at which banks make loans to one another. Whenever the Federal Open Market Committee meets, it chooses a target for the federal funds rate. The Fed's bond traders are then told to conduct open-market operations to hit the desired target.

The hard part of the Fed's job is choosing the target for the federal funds rate. Two general guidelines are clear. First, when inflation heats up, the federal funds rate should rise. An increase in the interest rate will mean a smaller money supply and, eventually, lower investment, lower output, higher unemployment, and reduced inflation. Second, when real economic activity slows—as reflected in real GDP or unemployment—the federal funds rate should fall. A decrease in the interest rate will mean a larger money supply and, eventually, higher investment, higher output, and lower unemployment. These two guidelines are represented by the monetary-policy equation in the dynamic *AD–AS* model.

The Fed needs to go beyond these general guidelines, however, and decide how much to respond to changes in inflation and real economic activity. Stanford University economist John Taylor has proposed the following rule for the federal funds rate:[1]

$$\text{Nominal Federal Funds Rate} = \text{Inflation} + 2.0 + 0.5\ (\text{Inflation} - 2.0) + 0.5\ (\text{GDP gap}).$$

$$\text{Nominal Federal Funds Rate} = \text{Inflation} + 2.0 + 0.5\ (\text{Inflation} - 2.0) + 0.5\ (\text{GDP gap}).$$

The *GDP gap* is the percentage by which real GDP deviates from an estimate of its natural level. (For consistency with our dynamic *AD–AS* model, the GDP gap here is taken to be positive if GDP rises above its natural level and negative if it falls below it.)

According to the **Taylor rule**, the real federal funds rate—the nominal rate minus inflation—should respond to inflation and the GDP gap. According to this rule, the real federal funds rate equals 2 percent when inflation is 2 percent and GDP is at its natural level. The first constant of 2 percent in this equation can be interpreted as an estimate of the natural rate of interest ρ, and the second constant of 2 percent subtracted from inflation can be interpreted as the Fed's inflation target π_t^*. For each percentage point that inflation rises above 2 percent, the real federal funds rate rises by 0.5 percent. For each percentage point that real GDP rises above its natural level, the real federal funds rate rises by 0.5 percent. If inflation falls below 2 percent or GDP moves below its natural level, the real federal funds rate falls accordingly.

In addition to being simple and reasonable, the Taylor rule for monetary policy also resembles actual Fed behavior in recent years. Figure 15-1 shows the actual nominal federal funds rate and the target rate as determined by Taylor's proposed rule. Notice how the two series tend to move together. John Taylor's monetary rule may be more than an academic suggestion. To some degree, it may be the rule that the Fed governors subconsciously follow.

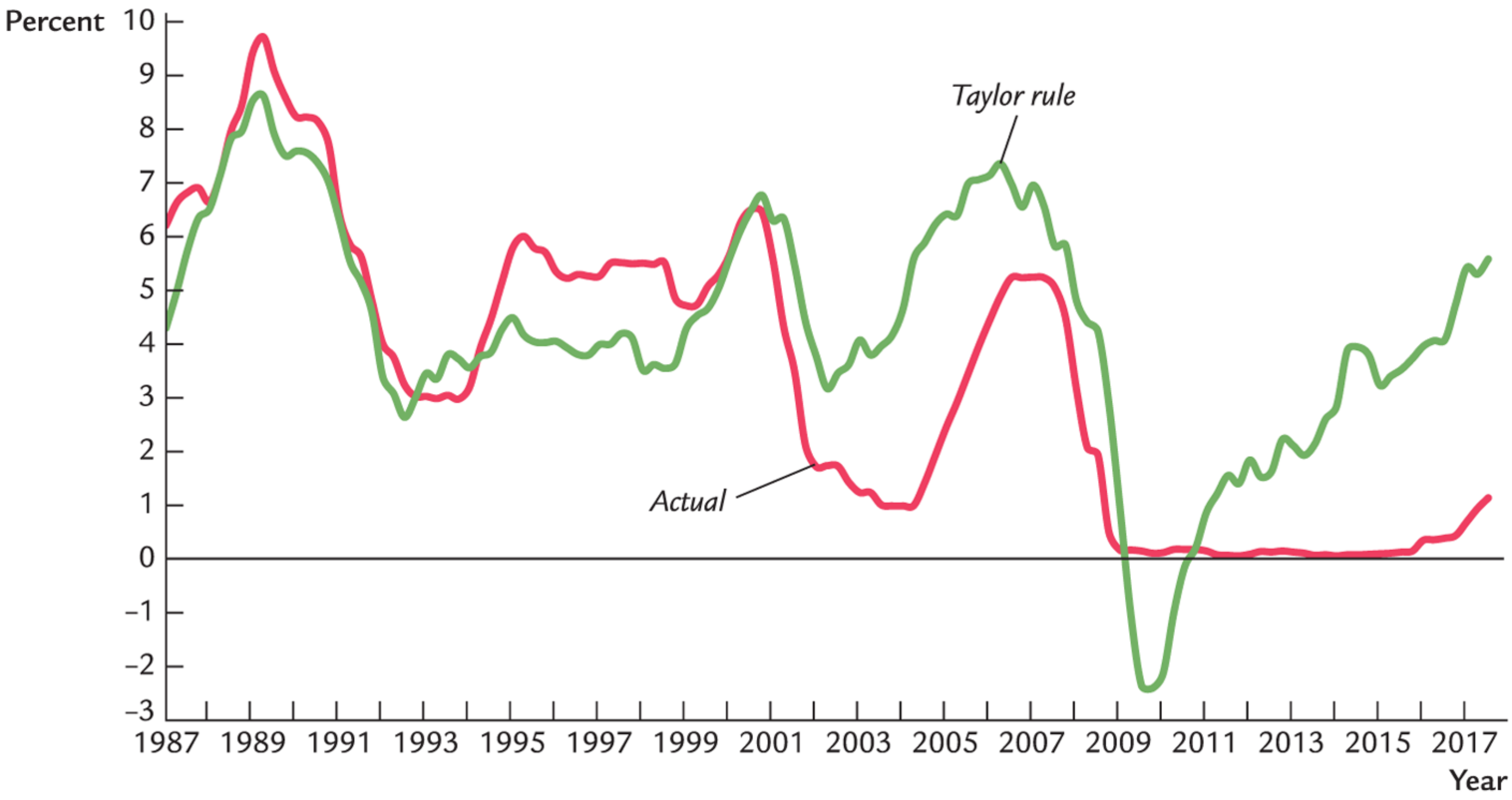


FIGURE 15-1 The Federal Funds Rate: Actual and Suggested This figure shows the federal funds rate set by the Federal Reserve and the target rate that the Taylor rule for monetary policy would recommend. Notice that the two series move closely together.

Data from: Federal Reserve Board, U.S. Department of Commerce, U.S. Department of Labor, and author's calculations. To implement the Taylor rule, the inflation rate is measured as the percentage change in the GDP deflator over the previous four quarters, and the GDP gap is measured as negative two times the deviation of the unemployment rate from its natural rate (as shown in Figure 7-1).

Notice that if inflation and output are both low enough, the Taylor rule can prescribe a negative nominal interest rate, as indeed occurred during the Great Recession of 2008–2009. Such a policy is not feasible, however. As we discussed in Chapter 12, a central bank cannot set a negative nominal interest rate because people would just hold currency (which pays a zero nominal return) rather than lend at a negative rate. In these circumstances, the Taylor rule cannot be strictly followed. The closest a central bank can come to following the rule is to set the interest rate at about zero, as the Fed did during this period.

The Taylor rule started recommending an increase in the federal funds rate around 2011. The Fed, however, kept the interest rates at about zero. This recent discrepancy has been a source of debate. Some economists argue that the Fed's policy was appropriate to make up for the period when interest rates were above the negative levels the rule advised. That is, they believe that to help the economy recover from the Great Recession, a period of below-rule interest rates was needed to compensate for the preceding period of above-rule interest rates. Other economists have suggested that the natural rate of interest had fallen, so the constant term in the Taylor rule needed to be reduced. Still other economists argue that the Fed was too slow to raise interest rates as the economy recovered from the recession. They fear that continued low interest rates might sow the seeds of future inflationary pressures. ■

15-2 Solving the Model

We have now looked at each piece of the dynamic *AD–AS* model. As a summary, Table 15-1 lists the equations, variables, and parameters in the model. The variables are grouped according to whether they are *endogenous* (to be determined by the model) or *exogenous* (taken as given by the model).

TABLE 15-1 The Equations, Variables, and Parameters in the Dynamic *AD–AS* Model

Equations	
$Y_t = \bar{Y}_t - \alpha(r_t - \rho) + \varepsilon_t$	The demand for goods and services
$r_t = i_t - E_t\pi_{t+1}$	The Fisher equation
$\pi_t = E_{t-1}\pi_t + \phi(Y_t - \bar{Y}_t) + \upsilon_t$	The Phillips curve
$E_t\pi_{t+1} = \pi_t$	Adaptive expectations
$i_t = \pi_t + \rho + \theta_\pi(\pi_t - \pi_t) + \theta_Y(Y_t - \bar{Y}_t)$	The monetary-policy rule
Endogenous Variables	
Y_t	Output
π_t	Inflation
r_t	Real interest rate
i_t	Nominal interest rate
$E_t\pi_{t+1}$	Expected inflation
Exogenous Variables	
$\bar{Y}_t$	Natural level of output
π_t	Central bank’s target for inflation
ε_t	Shock to the demand for goods and services
υ_t	Shock to the Phillips curve (supply shock)
Predetermined Variable	
π_{t-1}	Previous period’s inflation
Parameters	
α	The responsiveness of the demand for goods and services to the real interest

	rate
ρ	The natural rate of interest
ø	The responsiveness of inflation to output in the Phillips curve
θπ	The responsiveness of the nominal interest rate to inflation in the monetary-policy rule
θY	The responsiveness of the nominal interest rate to output in the monetary-policy rule

The model's five equations determine the paths of five endogenous variables: output Y_t, the real interest rate r_t, inflation π_t, expected inflation $E_t\pi_{t+1}$, and the nominal interest rate i_t. In any period, the five endogenous variables are influenced by the four exogenous variables in the equations as well as the previous period's inflation rate. Lagged inflation π_{t-1} is called a *predetermined variable*. That is, it is a variable that was endogenous in the past but, because it is fixed by the time when we arrive in period *t*, it is essentially exogenous for the purposes of finding the current equilibrium.

We are almost ready to put together these pieces to see how various shocks to the economy influence the paths of these variables over time. Before doing so, however, we need to establish the starting point for our analysis: the economy's long-run equilibrium.

The Long-Run Equilibrium

The long-run equilibrium represents the normal state around which the economy fluctuates. It occurs when there are no shocks $(\varepsilon_t = \nu_t = 0)$ and inflation has stabilized $(\pi_t = \pi_{t-1})$.

Straightforward algebra applied to the model's five equations can be used to determine the long-run values of the five endogenous variables:

$$\begin{aligned} Y_t &= \overline{Y}_t, \\ r_t &= \rho, \\ \pi_t &= \pi^*_t, \\ E_t\pi_{t+1} &= \pi^*_t, \\ i_t &= \rho + \pi^*_t. \end{aligned}$$

In words, the long-run equilibrium is described as follows: output and the real interest rate are at their natural values, inflation and expected inflation are at the target rate of inflation, and the nominal interest rate equals the natural rate of interest plus target inflation.

The long-run equilibrium of this model reflects two related principles: the classical dichotomy and monetary neutrality. Recall that the classical dichotomy is the separation of real from nominal variables and that monetary neutrality is the property according to which monetary policy does not influence real variables. The equations immediately above show that the central bank's inflation target π_t^* influences only inflation π_t, expected inflation $E_t\pi_{t+1}$, and the nominal interest rate i_t. If the central bank raises its inflation target, then inflation, expected inflation, and the nominal interest rate all increase by the same amount. Monetary policy does not influence the real variables—output Y_t and the real interest rate r_t. In these ways, the long-run equilibrium of the dynamic *AD–AS* model mirrors the classical models we examined in Chapters 3 to 9.

The Dynamic Aggregate Supply Curve

To study the behavior of this economy in the short run, it is useful to analyze the model graphically. Because graphs have two axes, we need to focus on two variables. We will use output Y_t and inflation π_t because these are the variables of central interest. As in the conventional *AD–AS* model, output will be on the horizontal axis. But because the price level has now faded into the background, the vertical axis in our graphs will now represent the inflation rate.

To generate this graph, we need two equations that summarize the relationships between output Y_t and inflation π_t. These equations are derived from the five equations of the model we have already seen. To isolate the relationships between Y_t and π_t, however, we need to use a bit of algebra to eliminate the other three endogenous variables $(r_t,\ i_t,$ and $E_t\pi_{t+1})$.

The first relationship between output and inflation comes almost directly from the Phillips curve equation. We can get rid of the one endogenous variable in the equation $(E_{t-1}\pi_t)$ by using the expectations equation $(E_{t-1}\pi_t = \pi_{t-1})$ to substitute past inflation π_{t-1} for expected inflation $E_{t-1}\pi_t$. With this substitution, the equation for the Phillips curve becomes

$$\pi_t = \pi_{t-1} + \phi(Y_t - \overline{Y}_t) + \nu_t. \qquad (DAS)$$

This equation relates inflation π_t and output Y_t for given values of two exogenous variables (natural output $\overline{Y}_t$ and a supply shock ν_t) and a predetermined variable (the previous period's inflation rate π_{t-1}).

[Figure 15-2](#) graphs the relationship between inflation π_t and output Y_t described by this equation. We call this upward-sloping curve the *dynamic aggregate supply* (or *DAS*) *curve*. The dynamic aggregate supply curve is similar to the aggregate supply curve from Chapter 14 except that inflation rather than the price level is on the vertical axis. The *DAS* curve shows how inflation is related to output in the short run. Its upward slope reflects the Phillips curve: other things equal, higher levels of economic activity are associated with higher marginal costs of production and, therefore, higher inflation.

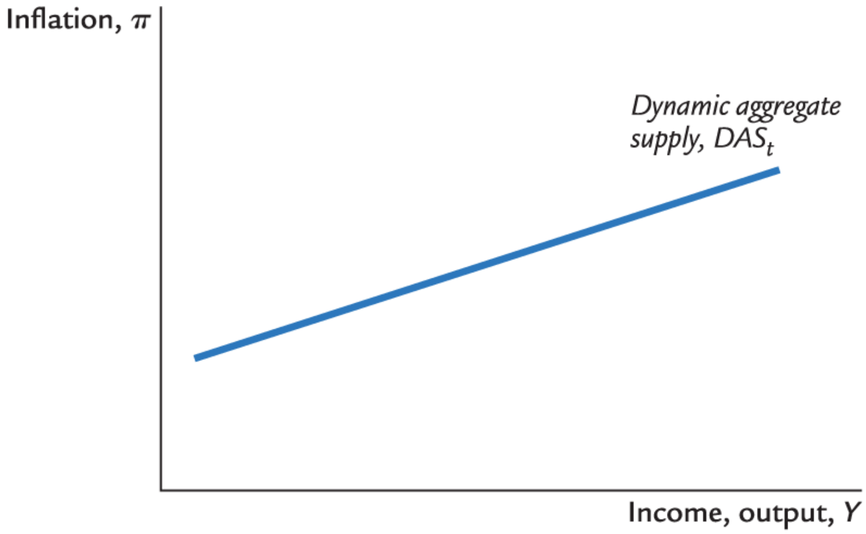


FIGURE 15-2 The Dynamic Aggregate Supply Curve The dynamic aggregate supply curve DAS_t shows a positive association between output Y_t and inflation π_t. Its upward slope reflects the Phillips curve relationship: other things equal, high levels of economic activity are associated with high inflation. The dynamic aggregate supply curve is drawn for given values of past inflation π_{t-1}, the natural level of output $\overline{Y}_t$, and the supply shock ν_t. When these variables change, the curve shifts.

The *DAS* curve is drawn for given values of past inflation π_{t-1}, the natural level of output $\overline{Y}_t$, and the supply shock ν_t. If any one of these three variables changes, the *DAS* curve shifts. One of our tasks ahead is to trace out the implications of such shifts. But first, we need another curve.

The Dynamic Aggregate Demand Curve

The dynamic aggregate supply curve is one of the two relationships between output and inflation that determine the economy’s short-run equilibrium. The other relationship is (no surprise) the dynamic aggregate demand curve. We derive it by combining four equations from the model and then eliminating all the endogenous variables other than output and inflation. Once we have an equation with only two endogenous variables Y_t and π_t), we can plot the relationship on our two-dimensional graph.

We begin with the demand for goods and services:

$$Y_t = \overline{Y}_t - \alpha\left(r_t - \rho\right) + \varepsilon_t.$$

To eliminate the endogenous variable r_t, the real interest rate, we use the Fisher equation to substitute $i_t - E_t\pi_{t+1}$ for r_t:

$$Y_t = \overline{Y}_t - \alpha\left(i_t - E_t\pi_{t+1} - \rho\right) + \varepsilon_t.$$

To eliminate another endogenous variable, the nominal interest rate i_t, we use the monetary-policy equation to substitute for i_t:

$$Y_t = \overline{Y}_t - \alpha\left[\pi_t + \rho + \theta_\pi\left(\pi_t - \pi*_t\right) + \theta_Y\left(Y_t - \overline{Y}_t\right) - E_t\pi_{t+1} - \rho\right] + \varepsilon_t.$$

Next, to eliminate the endogenous variable of expected inflation $E_t\pi_{t+1}$, we use our equation for inflation expectations to substitute π_t for $E_t\pi_{t+1}$:

$$Y_t = \overline{Y}_t - \alpha\left[\pi_t + \rho + \theta_\pi\left(\pi_t - \pi_t^*\right) + \theta_Y\left(Y_t - \overline{Y}_t\right) - \pi_t - \rho\right] + \varepsilon_t.$$

As was our goal, this equation has only two endogenous variables: output Y_t and inflation π_t. We can now simplify it. Notice that the positive π_t and ρ inside the brackets cancel the negative ones. The equation then becomes

$$Y_t = \overline{Y}_t - \alpha\left[\theta_\pi\left(\pi_t - \pi_t^*\right) + \theta_Y\left(Y_t - \overline{Y}_t\right)\right] + \varepsilon_t.$$

If we now bring like terms together and solve for Y_t, we obtain

$$Y_t = \overline{Y}_t - \left[\alpha\theta_\pi / \left(1 + \alpha\theta_Y\right)\right]\left(\pi_t - \pi_t^*\right) + \left[1/\left(1 + \alpha\theta_Y\right)\right]\varepsilon_t. \qquad (DAD)$$

This equation relates output Y_t to inflation π_t for given values of three exogenous variables $\left(\overline{Y}_t,\ \pi_t^*,\ \text{and } \varepsilon_t\right)$. In words, it says output equals the natural level of output when

(Y¯t, πt , and εt).

inflation is on target (πt=πt)$(\pi_t = \pi_t^*)$ and there is no demand shock (εt=0).$(\varepsilon_t = 0)$. Output rises above its natural level if inflation is below target (πt<πt)$(\pi_t < \pi_t^*)$ or if the demand shock is positive (εt>0). $(\varepsilon_t > 0)$. Output falls below its natural level if inflation is above target (πt>πt)$(\pi_t > \pi_t^*)$ or if the demand shock is negative (εt<0).$(\varepsilon_t < 0)$.

Figure 15-3 graphs the relationship between inflation πtπ_t and output YtY_t described by this equation. We call this downward-sloping curve the *dynamic aggregate demand* (or *DAD*) *curve*. The *DAD* curve shows how the quantity of output demanded is related to inflation in the short run. It is drawn holding constant the exogenous variables in the equation: the natural level of output Y¯t$\overline{Y}_t$, the inflation target πt ,π_t^*, and the demand shock εt.ε_t. If any one of these three exogenous variables changes, the *DAD* curve shifts. We will examine the effect of such shifts shortly.

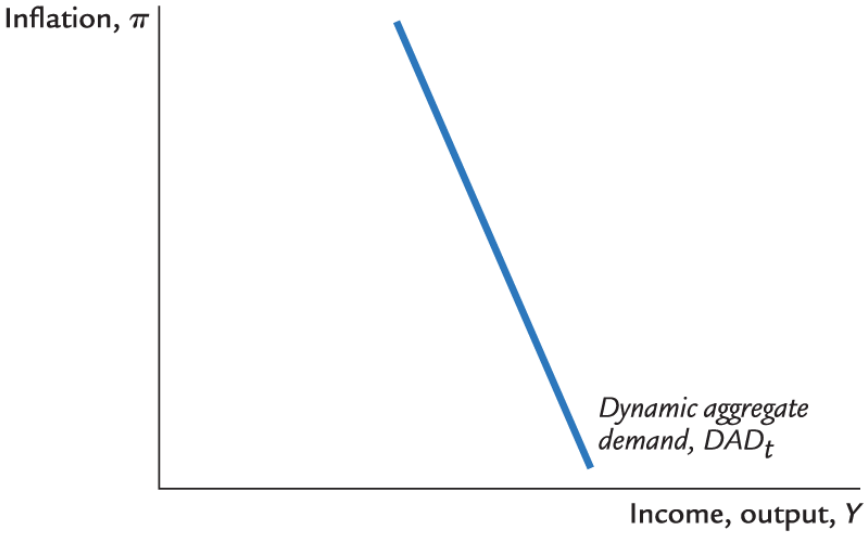


FIGURE 15-3 The Dynamic Aggregate Demand Curve The dynamic aggregate demand curve shows a negative association between output and inflation. Its downward slope reflects monetary policy and the demand for goods and services: a high level of inflation causes the central bank to raise nominal and real interest rates, which in turn reduces the demand for goods and services. The dynamic aggregate demand curve is drawn for given values of the natural level of output Y¯t,$\overline{Y}_t$, the inflation target πt ,π_t^*, and the demand shock εt.ε_t. When these exogenous variables change, the curve shifts.

It is tempting to think of this dynamic aggregate demand curve as just the standard aggregate demand curve from Chapter 12 with inflation, rather than the price level, on the vertical axis. In some ways, they are similar: both embody the link between the interest rate and the demand for goods and services. But there is an important difference. The conventional aggregate demand curve in Chapter 12 is drawn for a given money supply. By contrast, because the monetary-policy rule is used to derive the dynamic aggregate demand equation, the dynamic aggregate demand curve is drawn for a given rule for monetary policy. Under that rule, the central bank sets the interest rate based on macroeconomic conditions, and it allows the money supply to adjust accordingly.

The dynamic aggregate demand curve slopes downward because of the following mechanism. When inflation rises, the central bank responds by following its rule and increasing the nominal interest rate. Because the rule specifies that the central bank raise the nominal interest rate by more than the increase in inflation, the real interest rate rises as well. The increase in the real interest rate reduces the quantity of goods and services demanded. This negative association between inflation and quantity demanded, working through central bank policy, makes the dynamic aggregate demand curve slope downward.

The dynamic aggregate demand curve shifts in response to changes in fiscal and monetary policy. As we noted earlier, the shock variable ε_t reflects changes in government spending and taxes (among other things). Any change in fiscal policy that increases the demand for goods and services means a positive value of ε_t and a shift of the *DAD* curve to the right. Any change in fiscal policy that decreases the demand for goods and services means a negative value of ε_t and a shift of the *DAD* curve to the left.

Monetary policy enters the dynamic aggregate demand curve through the target inflation rate π^*_t. The *DAD* equation shows that, other things equal, an increase in π^*_t raises the quantity of output demanded. (There are two negative signs in front of π^*_t, so the effect is positive.) The mechanism behind this result is as follows. When the central bank raises its target for inflation, it pursues a more expansionary monetary policy by reducing the nominal interest rate, as dictated by the monetary-policy rule. For any given rate of inflation, the lower nominal interest rate results in a lower real interest rate, and a lower real interest rate stimulates spending on goods and services. Thus, output is higher for any given inflation rate, so the dynamic aggregate demand curve shifts to the right. Conversely, when the central bank reduces its target for inflation, it raises nominal and real interest rates, thereby dampening demand for goods and services and shifting the dynamic aggregate demand curve to the left.

The Short-Run Equilibrium

The economy's short-run equilibrium is determined by the intersection of the dynamic aggregate demand curve and the dynamic aggregate supply curve. The economy can be represented algebraically using the two equations we just derived:

Yt=Y¯t−[αθπ/(1+αθY)](πt−πt)+[1/(1+αθY)]εt,(DAD)πt=πt−1+ø(Yt−Y¯t)+νt.(DAS)

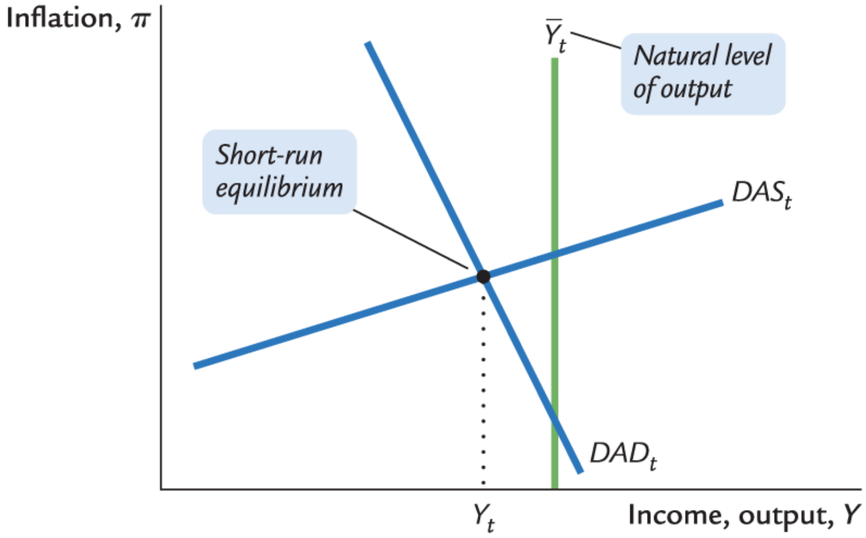
$$Y_t = \overline{Y}_t - [\alpha\theta_\pi / (1+\alpha\theta_Y)](\pi_t - \pi^*_t) + [1/(1+\alpha\theta_Y)]\varepsilon_t, \quad (DAD)$$

$$\pi_t = \pi_{t-1} + \phi(Y_t - \overline{Y}_t) + \nu_t. \quad (DAS)$$

In any period *t*, these equations together determine two endogenous variables: inflation π_t and output Y_t. The solution depends on five other variables that are exogenous (or at least determined prior to period *t*).

These exogenous (and predetermined) variables are the natural level of output $\bar{Y}_t$, the central bank's target inflation rate π^*_t, the shock to demand ε_t, the shock to supply ν_t, and the previous period's rate of inflation π_{t-1}.

Taking these exogenous variables as given, we can illustrate the economy's short-run equilibrium as the intersection of the dynamic aggregate demand curve and the dynamic aggregate supply curve, as in Figure 15-4. The short-run equilibrium level of output Y_t can be less than its natural level $\bar{Y}_t$, as it is in this figure, greater than its natural level, or equal to it. As we have seen, when the economy is in long-run equilibrium, output is at its natural level $(Y_t = \bar{Y}_t)$.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 15-4 The Short-Run Equilibrium The short-run equilibrium is determined by the intersection of the dynamic aggregate demand curve and the dynamic aggregate supply curve. This equilibrium determines the inflation rate and level of output in period *t*. In the equilibrium shown in this figure, the short-run equilibrium level of output Y_t falls short of the economy's natural level of output $\bar{Y}_t$.

The short-run equilibrium determines not only the level of output Y_t but also the inflation rate π_t. In the subsequent period $(t+1)$, this inflation rate will become the lagged inflation rate that influences the position of the dynamic aggregate supply curve. This connection between periods generates the dynamic patterns that we examine in the next section. That is, one period of time is linked to the next through expectations about inflation. A shock in period *t* affects inflation in period *t*, which in turn affects the inflation that people expect for period $t+1$. Expected inflation in period $t+1$ affects the position of the dynamic aggregate supply curve in that period, which in turn affects output and inflation in period $t+1$, which then affects expected inflation in period $t+2$, and so on.

These linkages of economic outcomes across time periods will become clear as we work through a series of examples.

15-3 Using the Model

Let's now use the dynamic *AD–AS* model to analyze how the economy responds to changes in the exogenous variables. The four exogenous variables in the model are the natural level of output $\bar{Y}_t$, the supply shock ν_t, the demand shock ε_t, and the central bank's inflation target π_t^*. To keep things simple, we assume that the economy always begins in long-run equilibrium and is then subject to a change in one of the exogenous variables. We also assume that the other exogenous variables are held constant.

Long-Run Growth

The economy's natural level of output $\bar{Y}_t$ grows over time because of population growth, capital accumulation, and technological progress, as discussed in Chapters 8 and 9. For our purposes here, we can take such growth as exogenous—that is, determined outside of this model. Figure 15-5 shows the effect of an exogenous increase in $\bar{Y}_t$. Because the natural level of output affects both the dynamic aggregate demand curve and the dynamic aggregate supply curve, both curves shift. In fact, they both shift to the right by exactly the amount that $\bar{Y}_t$ has increased.

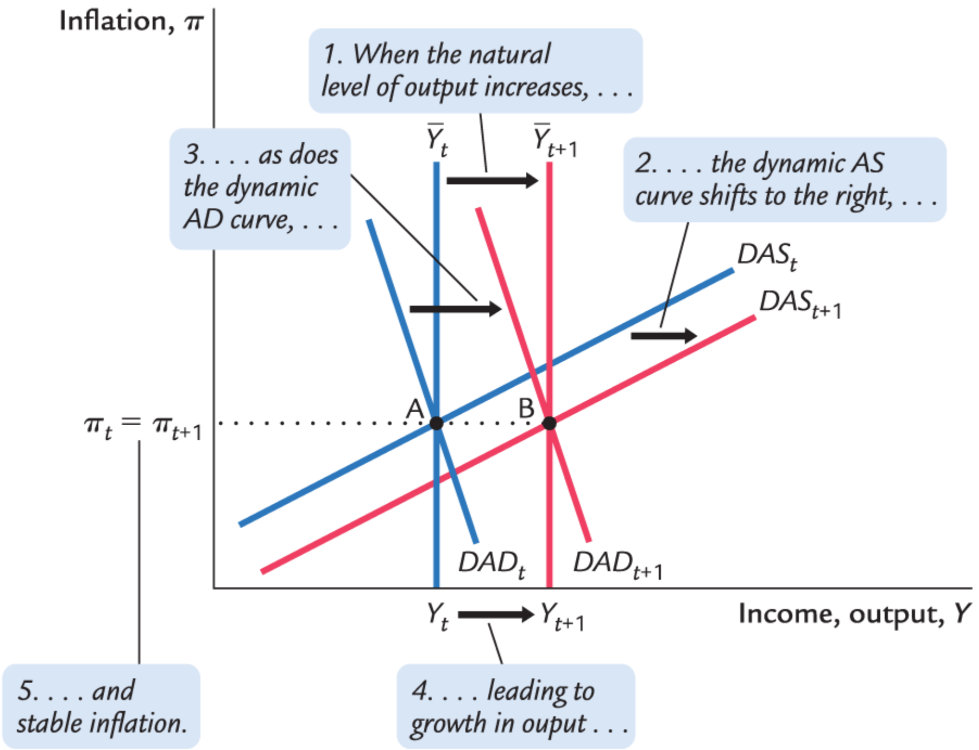


FIGURE 15-5 Long-Run Growth When long-run growth causes the natural level of output $\bar{Y}_t$ to increase, both the dynamic aggregate demand curve and the dynamic aggregate supply curve shift to the right by the same amount. Output Y_t increases, but inflation π_t remains the same.

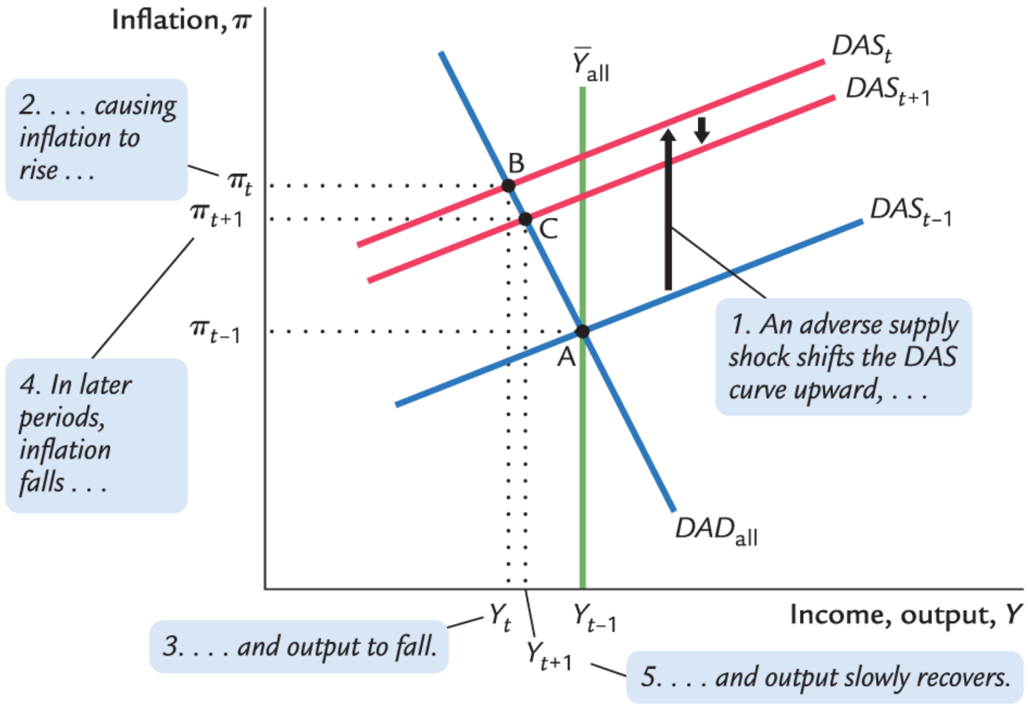
The shifts in these curves move the economy's equilibrium in the figure from point A to point B. Output Y_t increases by as much as the natural level $\overline{Y}_t$. Inflation is unchanged.

The story behind these conclusions is as follows. When the natural level of output increases, the economy can produce a larger quantity of goods and services. This is represented by the rightward shift in the dynamic aggregate supply curve. At the same time, the increase in the natural level of output makes people richer. Other things equal, they want to buy more goods and services. This is represented by the rightward shift in the dynamic aggregate demand curve. The simultaneous shifts in supply and demand increase the economy's output without putting either upward or downward pressure on inflation. In this way, the economy can experience long-run growth and a stable inflation rate.

A Shock to Aggregate Supply

Now consider a shock to aggregate supply. Suppose ν_t rises to 1 percent for one period and subsequently returns to zero. This shock to the Phillips curve might occur, for example, because turmoil in the Middle East pushes up oil prices or because a drought drives up food prices. In general, the supply shock ν_t captures any event that influences inflation other than expected inflation $E_{t-1}\pi_t$ and current economic activity, as measured by $Y_t - \overline{Y}_t$.

Figure 15-6 shows the result. In period *t*, when the shock occurs, the dynamic aggregate supply curve shifts upward from DAS_{t-1} to DAS_t. To be precise, the curve shifts upward by the size of the shock, which we assumed to be 1 percentage point. Because the supply shock ν_t is not a variable in the dynamic aggregate demand equation, the *DAD* curve is unchanged. Therefore, the economy moves along the dynamic aggregate demand curve from point A to point B. As the figure illustrates, the supply shock in period *t* causes inflation to rise to π_t and output to fall to Y_t.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 15-6 A Supply Shock A supply shock in period *t* shifts the dynamic aggregate supply curve upward from DASt−1 DAS_{t-1} to DASt. DAS_t. The dynamic aggregate demand curve is unchanged. The economy's short-run equilibrium moves from point A to point B. Inflation rises and output falls. In the subsequent period (t+1), $(t+1)$, the dynamic aggregate supply curve shifts to DASt+1, DAS_{t+1}, and the economy moves to point C. The supply shock has returned to its normal value of zero, but inflation expectations remain high. As a result, the economy returns only gradually to its initial equilibrium, point A.

These effects work in part through the reaction of monetary policy to the shock. When the supply shock increases inflation, the central bank responds by following its policy rule and raising nominal and real interest rates. The higher real interest rate reduces the quantity of goods and services demanded, which depresses output below its natural level. (This series of events is represented by the movement along the *DAD* curve from point A to point B.) The lower level of output dampens the inflationary pressure to some degree, so inflation rises somewhat less than the initial shock.

In the periods after the shock, expected inflation is higher because expectations depend on past inflation. In period t+1, $t+1$, for instance, the economy is at point C. Even though the shock variable νt ν_t returns to its normal value of zero, the dynamic aggregate supply curve does not immediately return to its initial position. Instead, it slowly shifts back downward toward its initial position DASt−1 DAS_{t-1} as a lower level of economic activity reduces inflation and thereby expectations of future inflation. Eventually, the economy is back at point A. Throughout the transition process, however, output remains below its natural level.

As the economy responds to the supply shock by moving in Figure 15-6 from point A to B to C and then gradually back to point A, all the variables in the model respond accordingly. Figure 15-7 shows the time paths of the key variables. (These simulations are based on realistic parameter values: see the nearby FYI box for their description.) As panel (a) shows, the shock νt ν_t spikes upward by 1 percentage point in period *t* and then returns to zero in subsequent periods. Inflation, shown in panel (d), rises by 0.9 percentage point and slowly returns to its target of 2 percent over a long period of time. Output, shown in panel (b), falls in response to the supply shock but also gradually returns to its natural level.

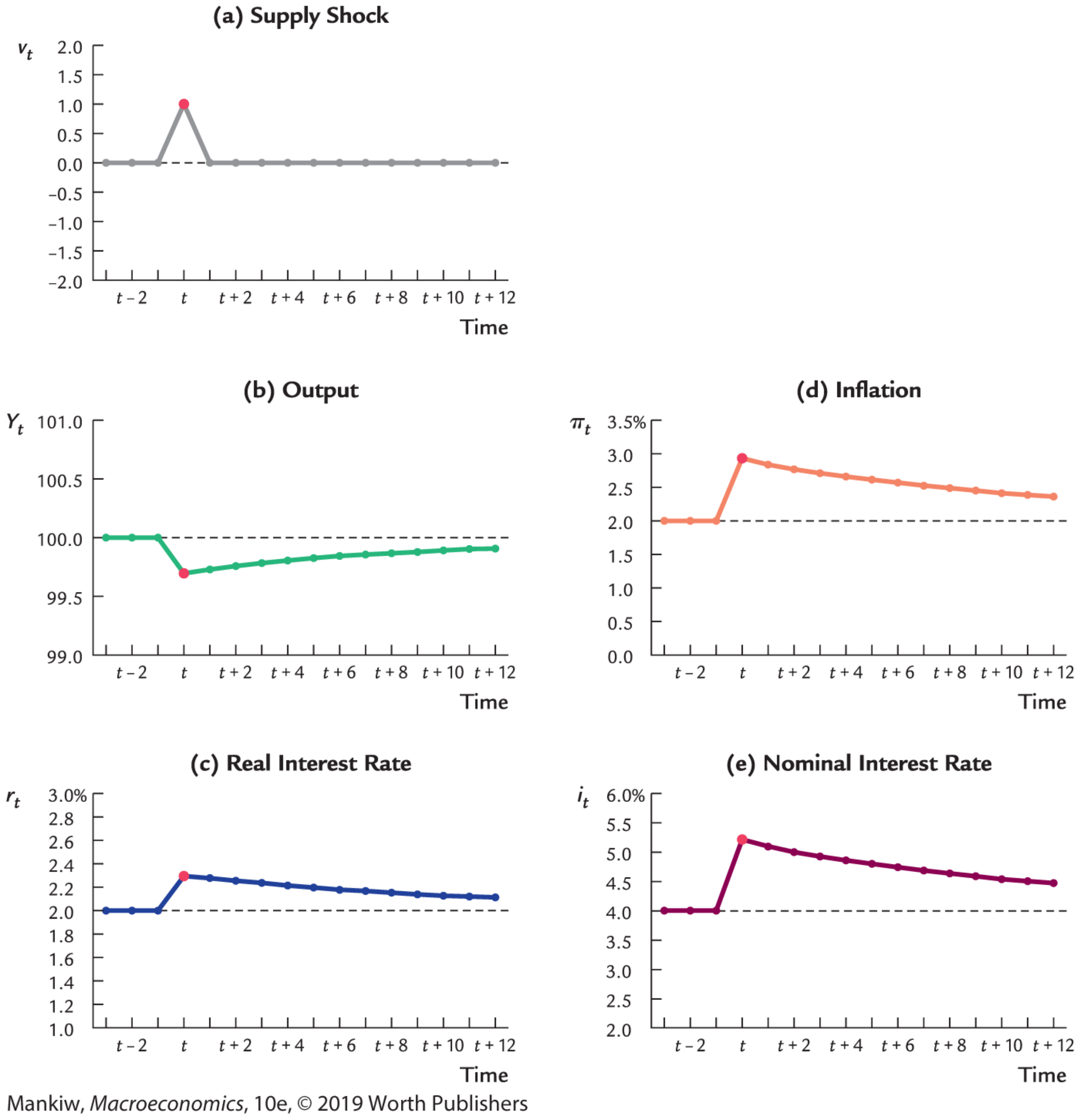


FIGURE 15-7 The Dynamic Response to a Supply Shock This figure shows the responses of the key variables over time to a one-time supply shock.

Figure 15-7 also shows the paths of nominal and real interest rates. In the period of the supply shock, the nominal interest rate, shown in panel (e), increases by 1.2 percentage points, and the real interest rate, in panel (c), increases by 0.3 percentage point. Both interest rates return to their normal values as the economy returns to its long-run equilibrium.

These figures illustrate the phenomenon of *stagflation* in the dynamic *AD–AS* model. A supply shock causes inflation to rise, which in turn increases expected inflation. As the central bank applies its rule for monetary policy and responds by raising interest rates, it gradually squeezes inflation out of the system, but only at the cost of a prolonged downturn in economic activity.

FYI

The Numerical Calibration and Simulation

The text presents some numerical simulations of the dynamic *AD–AS* model. When interpreting these results, it is easiest to think of each period as representing one year. We examine the impact of the change in the year of the shock (period *t*) and over the subsequent 12 years.

The simulations use these parameter values:

$$\begin{aligned} \bar{Y}_t &= 100, \\ \pi_t^* &= 2.0, \\ \alpha &= 1.0, \\ \rho &= 2.0, \\ \phi &= 0.25, \\ \theta_\pi &= 0.5, \\ \theta_Y &= 0.5. \end{aligned}$$

Here is how to interpret these numbers. The natural level of output $\bar{Y}_t$ is 100; by choosing this convenient number, we can view fluctuations in $Y_t - \bar{Y}_t$ as percentage deviations of output from its natural level. The central bank's inflation target π_t^* is 2 percent. The parameter $\alpha = 1.0$ implies that a 1-percentage-point increase in the real interest rate reduces output demand by 1, which is 1 percent of its natural level. The economy's natural rate of interest ρ is 2 percent. The Phillips curve parameter $\phi = 0.25$ implies that when output is 1 percent above its natural level, inflation rises by 0.25 percentage point. The parameters for the monetary policy rule $\theta_\pi = 0.5$ and $\theta_Y = 0.5$ are those suggested by John Taylor and are reasonable approximations of the Fed's behavior.

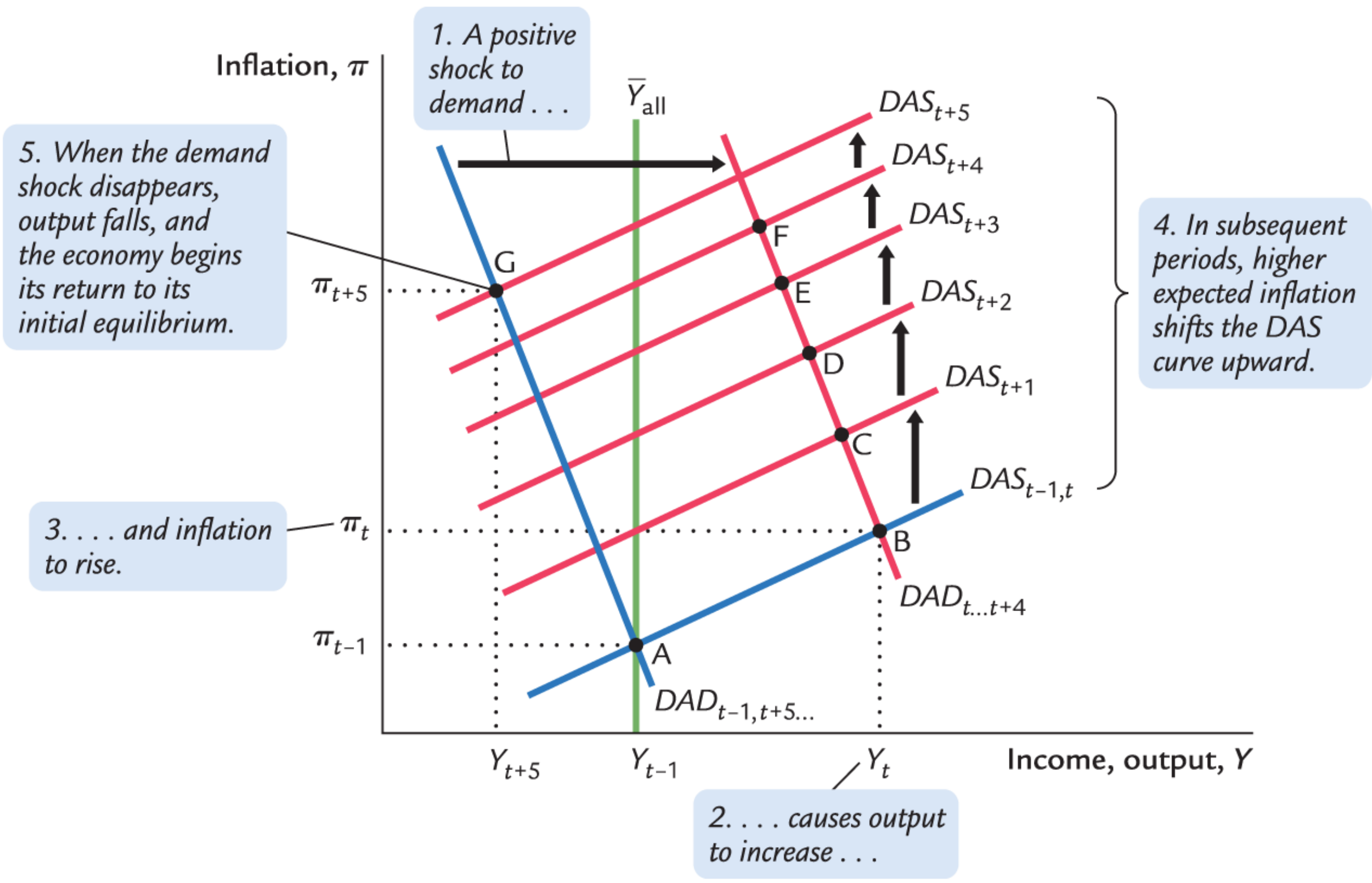
In all cases, the simulations assume a change of 1 percentage point in the exogenous variable of interest. Larger shocks would have qualitatively similar effects, but the magnitudes would be proportionately greater. For example, a shock of 3 percentage points would affect all the variables in the same way as a shock of 1 percentage point, but the movements would be three times as large as those in the simulation shown.

The graphs of the time paths of the variables after a shock (shown in Figures 15-7, 15-9, and 15-11) are called *impulse response functions*. The word *impulse* here refers to the shock, and *response function* refers to how the endogenous variables respond to the shock over time. These simulated impulse response functions provide one way to illustrate how the model works. They show how the endogenous variables move when a shock hits the economy, how these variables adjust in subsequent periods, and how they are correlated with one another over time.

A Shock to Aggregate Demand

Now let’s consider a shock to aggregate demand. To be realistic, the shock is assumed to persist over several periods. In particular, suppose $\varepsilon_t = 1$ for five periods and then returns to its normal value of zero. This positive shock ε_t might represent, for example, a war that increases government purchases or a stock market bubble that increases wealth and thereby consumption spending. In general, the demand shock captures any event that influences the demand for goods and services for given values of the natural level of output $\overline{Y}_t$ and the real interest rate r_t.

Figure 15-8 shows the result. In period *t*, when the shock occurs, the dynamic aggregate demand curve shifts to the right from DAD_{t-1} to DAD_t. Because the demand shock ε_t is not a variable in the dynamic aggregate supply equation, the *DAS* curve is unchanged from period $t-1$ to period *t*. The economy moves along the dynamic aggregate supply curve from point A to point B. Output and inflation both increase.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 15-8 A Demand Shock This figure shows the effects of a positive demand shock in period *t* that lasts for five periods. The shock immediately shifts the dynamic aggregate demand curve to the right from DAD_{t-1} to DAD_t. The economy moves from point A to point B. Both inflation and output rise. In the next period, the dynamic aggregate supply curve shifts to DAS_{t+1} because of increased expected inflation. The economy moves from point B to point C and then, in subsequent periods, to points D, E, and F. When the demand shock disappears after five periods, the dynamic aggregate demand curve shifts back to its initial position, and the economy moves from point F to point G. Output falls below its natural level, and inflation starts to fall. Over time, the dynamic aggregate supply curve starts shifting downward, and the economy gradually returns to its initial equilibrium, point A.

Once again, these effects work in part through the reaction of monetary policy to the shock. When the demand shock causes output and inflation to rise, the central bank responds by increasing the nominal and real interest rates. Because a higher real interest rate reduces the quantity of goods and services demanded, it partly offsets the expansionary effects of the demand shock.

In the periods after the shock occurs, expected inflation is higher because expectations depend on past inflation. As a result, the dynamic aggregate supply curve shifts upward repeatedly; as it does so, it continually reduces output and increases inflation. In the figure, the economy goes from point B in the initial period of the shock to points C, D, E, and F in subsequent periods.

In the sixth period (t+5), $(t + 5)$, the demand shock disappears. At this time, the dynamic aggregate demand curve returns to its initial position. However, the economy does not immediately return to its initial equilibrium, point A. The period of high demand has increased inflation and thereby expected inflation. High expected inflation keeps the dynamic aggregate supply curve higher than it was initially. As a result, when demand falls off, the economy's equilibrium moves to point G, and output falls to Yt+5, Y_{t+5}, which is below its natural level. The economy then gradually recovers, as the low level of output squeezes the higher-than-target inflation out of the system. Over time, as inflation and expected inflation fall, the economy slowly returns to point A.

Figure 15-9 shows the time path of the key variables in the model in response to the demand shock. Note that the positive demand shock increases real and nominal interest rates. When the demand shock disappears, both interest rates fall. These responses occur because when the central bank sets the nominal interest rate, it takes into account both inflation rates and deviations of output from its natural level.

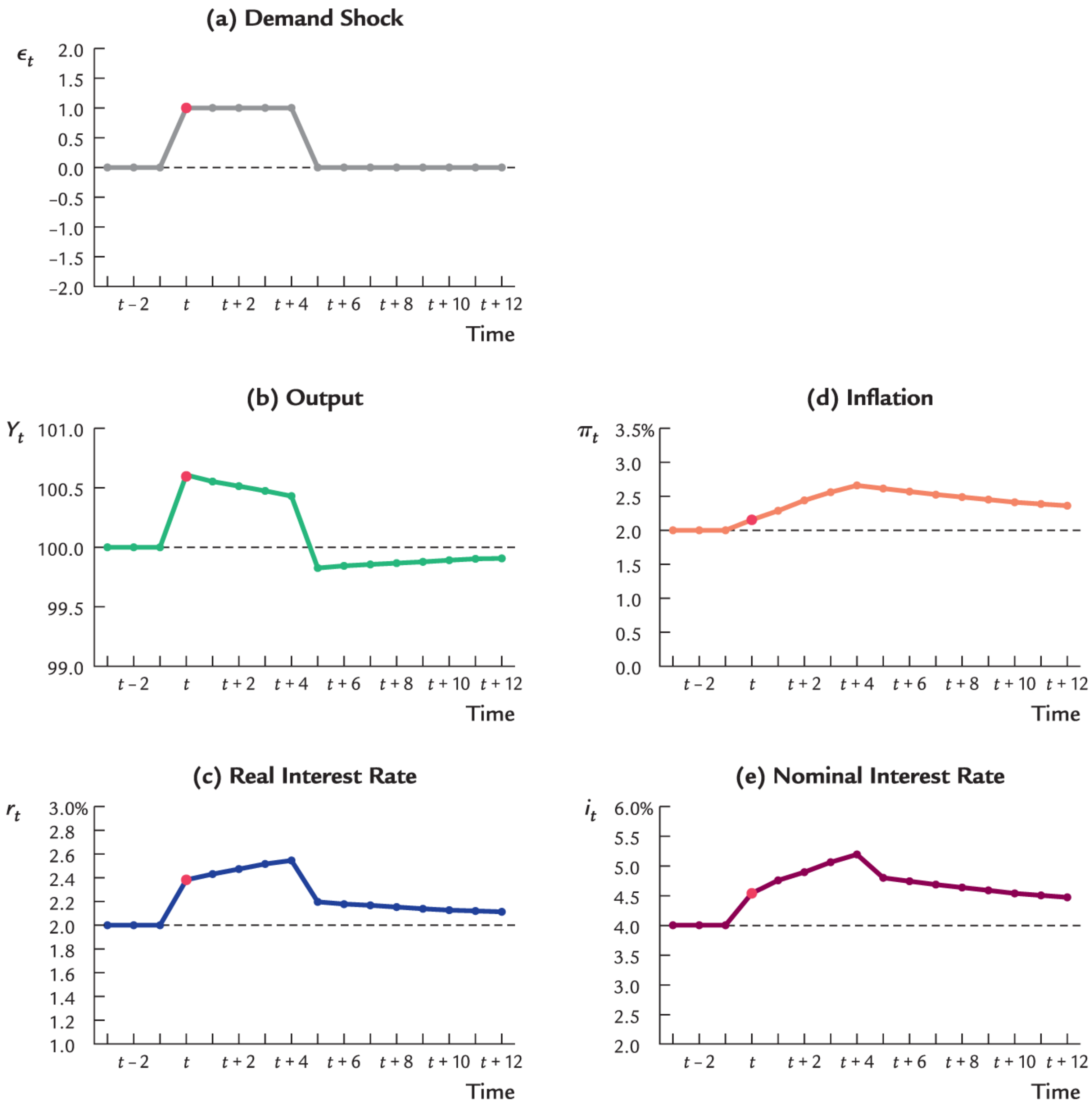
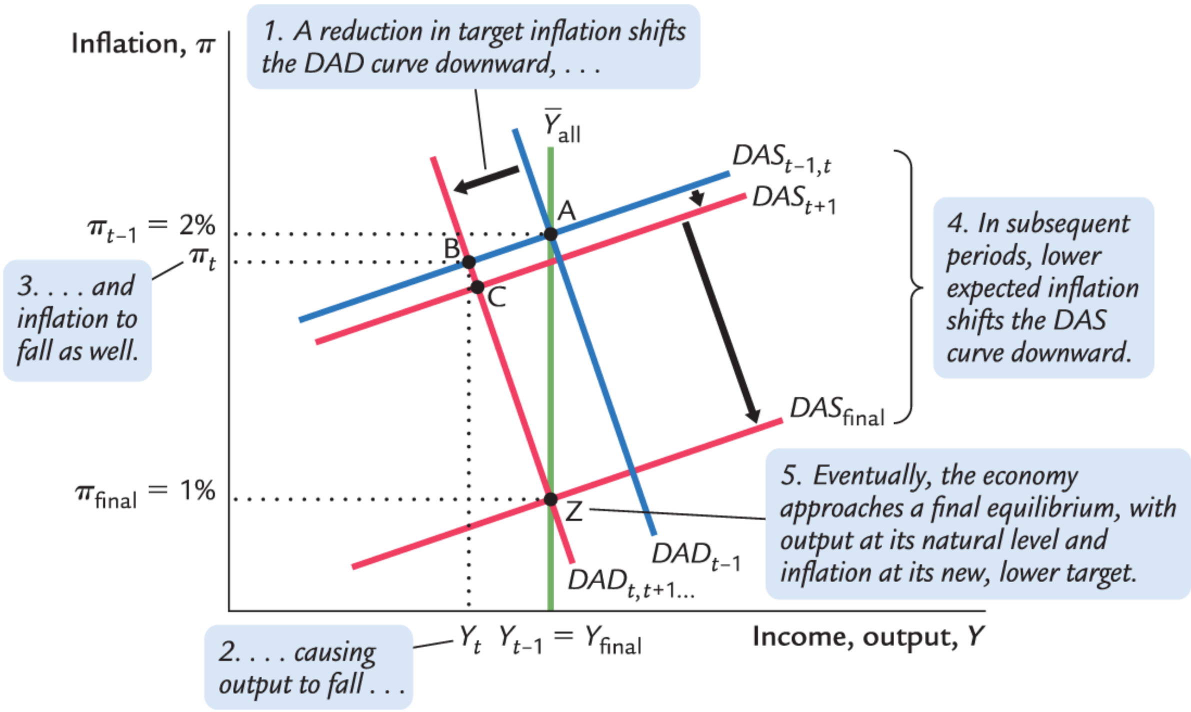


FIGURE 15-9 The Dynamic Response to a Demand Shock This figure shows the responses of the key variables over time to a positive 1-percent demand shock that lasts five periods.

A Shift in Monetary Policy

Suppose the central bank decides to reduce its target for the inflation rate. Specifically, imagine that, in period t, πt π_t^* falls from 2 percent to 1 percent and thereafter remains at that lower level. Let's consider how the economy will react to this change in monetary policy.

Recall that the inflation target enters the model as an exogenous variable in the dynamic aggregate demand curve. When the inflation target falls, the *DAD* curve shifts to the left, as shown in [Figure 15-10](). (To be precise, it shifts downward by 1 percentage point.) Because target inflation does not enter the dynamic aggregate supply equation, the *DAS* curve does not shift initially. The economy moves from its initial equilibrium, point A, to a new equilibrium, point B. Output falls below its natural level. Inflation falls as well, but not by the full 1 percentage point by which the central bank has lowered its inflation target.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 15-10 A Reduction in Target Inflation A permanent reduction in target inflation in period *t* shifts the dynamic aggregate demand curve to the left from DADt−1 DAD_{t-1} to DADt, DAD_t, where it then stays. Initially, the economy moves from point A to point B. Both inflation and output fall. In the subsequent period, because expected inflation falls, the dynamic aggregate supply curve shifts downward. The economy moves from point B to point C in period t+1. $t+1$. Over time, as expected inflation falls and the dynamic aggregate supply curve repeatedly shifts downward, the economy approaches a new equilibrium at point Z. Output returns to its natural level Y¯all, $\bar{Y}_{all}$, and inflation ends at its new, lower target (1 percent).

Monetary policy is key to the explanation of this outcome. Because the central bank has just lowered its target for inflation, current inflation is running above the new target. The central bank reacts by following its policy rule and raising real and nominal interest rates. The higher real interest rate reduces the demand for goods and services. The Phillips curve tells us that when output falls, inflation falls as well.

Lower inflation, in turn, reduces the inflation rate that people expect to prevail in the next period. In period t+1, $t+1$, lower expected inflation shifts the dynamic aggregate supply curve downward, to DASt+1. DAS_{t+1}. (To be precise, the curve shifts downward by exactly the fall in expected inflation.) This shift moves the economy from point B to point C, further reducing inflation and expanding output. Over time, as inflation continues to fall toward the new 1 percent target and the *DAS* curve continues to shift toward DASfinal, DAS_{final}, the economy approaches a new long-run equilibrium at point Z, where output is back at its natural level $(Y_{\text{final}} = \overline{Y}_{\text{all}})$ and inflation is at its new lower target $(\pi_{\text{final}} = 1 \text{ percent})$.

Figure 15-11 shows how the variables respond over time to a reduction in target inflation. Note in panel (e) the time path of the nominal interest rate i_t. Before the change in policy, the nominal interest rate is at its long-run value of 4.0 percent (which equals the natural real interest rate ρ of 2 percent plus target inflation π^*_{t-1} of 2 percent). When target inflation falls to 1 percent, the nominal interest rate rises to 4.2 percent. Over time, however, the nominal interest rate falls as inflation and expected inflation fall toward the new target rate; eventually, i_t approaches its new long-run value of 3.0 percent. Thus, a shift toward a lower inflation target increases the nominal interest rate in the short run but decreases it in the long run.

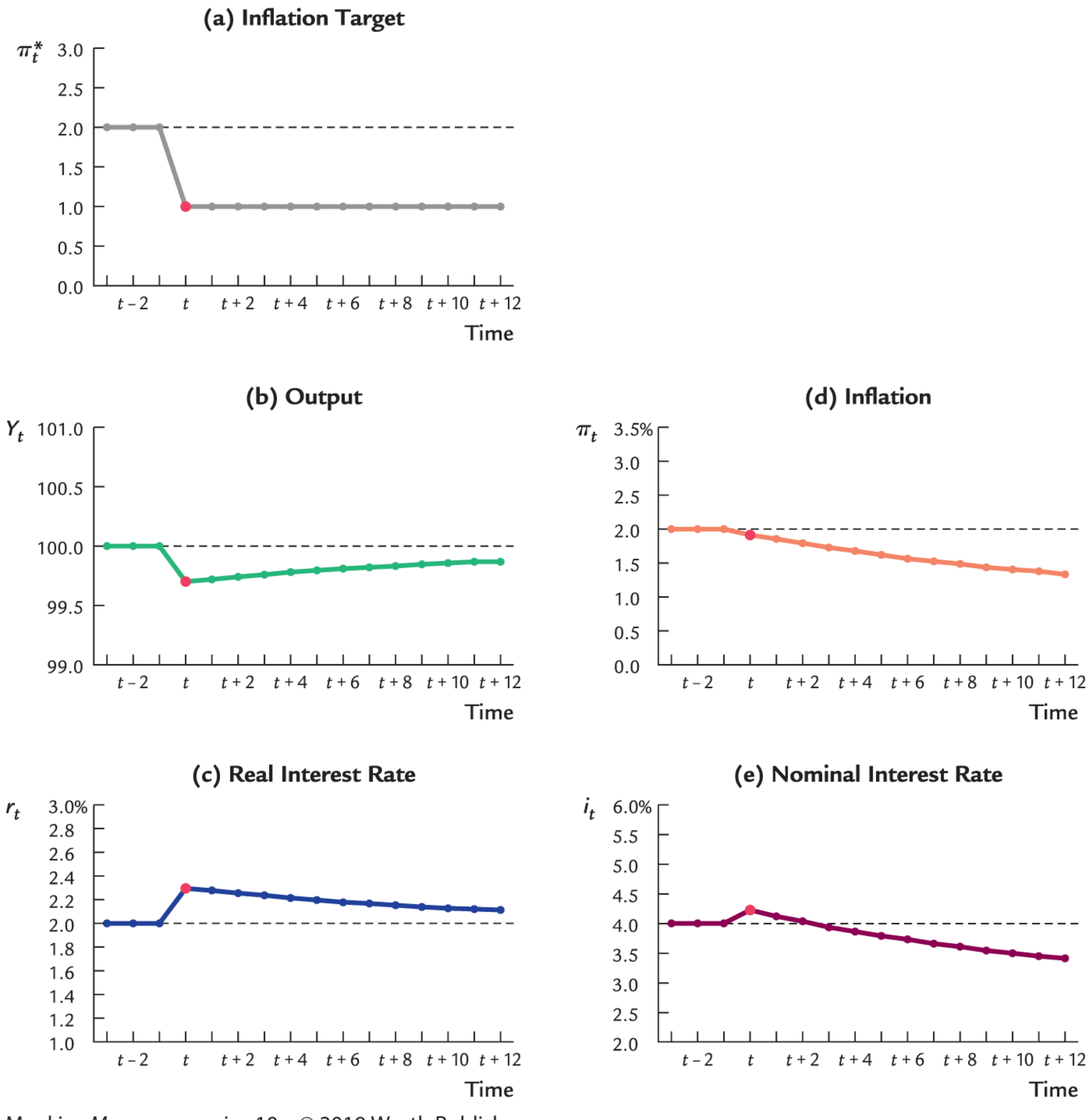


FIGURE 15-11 The Dynamic Response to a Reduction in Target Inflation This figure shows the responses of the key variables over time to a permanent reduction in the target rate of inflation.

We close with a caveat. Throughout this analysis we have maintained the assumption of adaptive expectations. That is, we have assumed that people form their expectations of inflation based on the inflation they have recently experienced. It is possible, however, that if the central bank makes a credible announcement of its new policy of lower target inflation, people will respond by immediately altering their expectations of inflation. That is, they may form expectations rationally, based on the policy announcement, rather than adaptively, based on what they have experienced. (We discussed this possibility in Chapter 14.) If so, the dynamic aggregate supply curve will shift downward immediately upon the change in policy, just when the dynamic aggregate demand curve shifts downward. In this case, the economy will instantly reach its new long-run equilibrium. By contrast, if people do not believe an announced policy of low inflation until they see it, then the assumption of adaptive expectations is appropriate, and the transition path to lower inflation will involve a period of lost output, as shown in Figure 15-11.

15-4 Two Applications: Lessons for Monetary Policy

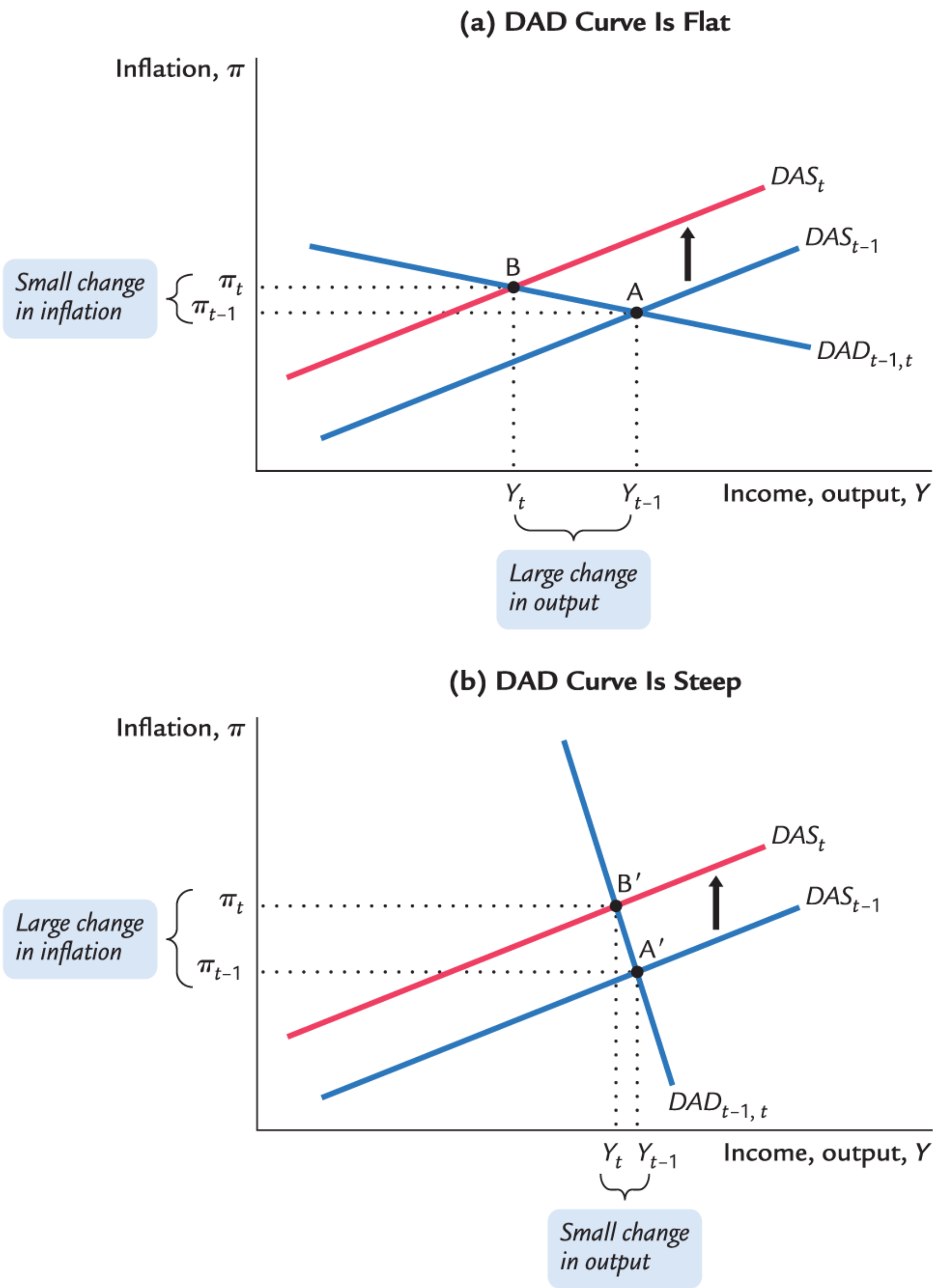
So far in this chapter, we have assembled a dynamic model of inflation and output and used it to show how various shocks affect the time paths of output, inflation, and interest rates. We now use the model to shed light on the design of monetary policy.

It is worth pausing at this point to consider what we mean by "the design of monetary policy." So far in this analysis, the central bank has had a simple role: it has merely had to adjust the money supply to ensure that the nominal interest rate hits the target level prescribed by the monetary-policy rule. The two key parameters of that policy rule are θ_π (the responsiveness of the target interest rate to inflation) and θ_Y (the responsiveness of the target interest rate to output). We have taken these parameters as given without discussing how they are chosen. Now that we know how the model works, we can consider a deeper question: What should the parameters of the monetary policy rule be?

The Tradeoff Between Output Variability and Inflation Variability

Consider the impact of a supply shock on output and inflation. According to the dynamic *AD–AS* model, the impact of this shock depends crucially on the slope of the dynamic aggregate demand curve. In particular, the slope of the *DAD* curve determines whether a supply shock has a large or small impact on output and inflation.

This phenomenon is illustrated in Figure 15-12. In the two panels of this figure, the economy experiences the same supply shock. In panel (a), the dynamic aggregate demand curve is nearly flat, so the shock has a small effect on inflation but a large effect on output. In panel (b), the dynamic aggregate demand curve is steep, so the shock has a large effect on inflation but a small effect on output.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 15-12 Two Possible Responses to a Supply Shock When the dynamic aggregate demand curve is relatively flat, as in panel (a), a supply shock has a small effect on inflation but a large effect on output. When the dynamic aggregate demand curve is relatively steep, as in panel (b), the same supply shock has a large effect on inflation but a small effect on output. The slope of the dynamic aggregate demand curve is based in part on the parameters of monetary policy θπθ_π and θYθ_Y), which describe how much interest rates respond to changes in inflation and output. When choosing these parameters, the central bank faces a tradeoff between the variability of inflation and the variability of output.

Why is this important for monetary policy? Because the central bank can influence the slope of the dynamic aggregate demand curve. Recall the equation for the *DAD* curve:

$$Y_t=\bar{Y}_t-[\ \alpha\theta_\pi/(1+\alpha\theta_Y)\ ](\pi_t-\pi_t\quad)+[\ 1/(1+\alpha\theta_Y)\ ]\varepsilon_t.$$

$$Y_t = \overline{Y}_t - [\alpha\theta_\pi / (1 + \alpha\theta_Y)] (\pi_t - \pi_t^*) + [1/ (1 + \alpha\theta_Y)] \varepsilon_t.$$

Two key parameters here are θ_π and θ_Y, which govern how much the central bank's interest rate target responds to inflation and output. When the central bank chooses these policy parameters, it determines the slope of the *DAD* curve and thus the economy's short-run response to supply shocks.

On the one hand, suppose the central bank responds strongly to inflation (θ_π is large) and weakly to output (θ_Y is small). In this case, the coefficient on inflation in the above equation is large. That is, a small change in inflation has a large effect on output. As a result, the dynamic aggregate demand curve is relatively flat, and supply shocks have large effects on output but small effects on inflation. The story goes like this. When the economy experiences a supply shock that pushes up inflation, the central bank's policy rule has it respond vigorously with higher interest rates. Sharply higher interest rates significantly reduce the quantity of goods and services demanded, thereby leading to a large recession that dampens the inflationary impact of the shock (which was the purpose of the monetary policy response).

On the other hand, suppose the central bank responds weakly to inflation (θ_π is small) but strongly to output (θ_Y is large). In this case, the coefficient on inflation in the above equation is small, meaning that even a large change in inflation has only a small effect on output. As a result, the dynamic aggregate demand curve is relatively steep, and supply shocks have small effects on output but large effects on inflation. The story is just the opposite of the story before. Now, when the economy experiences a supply shock that pushes up inflation, the central bank's policy rule has it respond with only slightly higher interest rates. This small policy response avoids a large recession but accommodates the inflationary shock.

In its choice of monetary policy, the central bank determines which of these two scenarios will play out. That is, when setting the policy parameters θ_π and θ_Y, the central bank chooses whether to make the economy look more like panel (a) or more like panel (b) of Figure 15-12. When making this choice, the central bank faces a tradeoff between output variability and inflation variability. The central bank can be a hard-line inflation fighter, as in panel (a), in which case inflation is stable but output is volatile. Alternatively, it can be more accommodative, as in panel (b), in which case inflation is volatile but output is more stable. It can also choose some position in between these two extremes.

One job of a central bank is to promote economic stability. There are, however, various dimensions to this goal. When there are tradeoffs to be made, the central bank has to determine what kind of stability to pursue. The dynamic *AD–AS* model shows that one fundamental tradeoff is between the variability in inflation and the variability in output.

Note that this tradeoff is very different from a simple tradeoff between inflation and output. In the long run of this model, inflation goes to its target, and output goes to its natural level. Consistent with classical macroeconomic theory, policymakers do not face a long-run tradeoff between inflation and output. Instead, they face a choice about which of these two measures of macroeconomic performance they want to stabilize. When deciding on the parameters of the monetary-policy rule, they determine whether supply shocks lead to inflation variability, output variability, or some combination of the two.

CASE STUDY

Different Mandates, Different Realities: The Fed Versus the ECB

According to the dynamic *AD–AS* model, a key policy choice facing any central bank concerns the parameters of its policy rule. The monetary parameters θ_π and θ_Y determine how much the interest rate responds to macroeconomic conditions. As we have just seen, these responses in turn determine the volatility of inflation and output.

The U.S. Federal Reserve and the European Central Bank (ECB) appear to have different approaches to this decision. The legislation that created the Fed states explicitly that its goal is “to promote effectively the goals of maximum employment, stable prices, and moderate long-term interest rates.” Because the Fed is supposed to stabilize both employment and prices, it is said to have a *dual mandate*. (The third goal—moderate long-term interest rates—should follow naturally from stable prices.) By contrast, the ECB says on its website that “the primary objective of the ECB’s monetary policy is to maintain price stability. The ECB aims at inflation rates of below, but close to, 2% over the medium term.” All other macroeconomic goals, including stability of output and employment, appear to be secondary.

We can interpret these differences in light of our model. Compared to the Fed, the ECB seems to give more weight to inflation stability and less weight to output stability. This difference in objectives should be reflected in the parameters of the monetary-policy rules. To achieve its dual mandate, the Fed would respond more to output and less to inflation than would the ECB.

The financial crisis of 2008–2009 illustrates these differences. In 2008, the world economy was experiencing rising oil prices, a financial crisis, and a slowdown in economic activity. The Fed responded to these events by reducing its target interest rate from 4.25 percent at the beginning of the year to a range of 0 to 0.25 percent at year’s end. The ECB, facing a similar situation, also cut interest rates but by much less—from 3 percent to 2 percent. It cut the interest rate to 0.25 percent only in 2009, when the depth of the recession was clear and inflationary worries had subsided. Throughout this episode, the ECB was less concerned about recession and more concerned about keeping inflation in check.

Although the dynamic *AD–AS* model predicts that, other things equal, the policy of the ECB should lead to more variable output and more stable inflation, testing this prediction is difficult. In practice, other things are rarely equal. Europe and the United States differ in many ways beyond the policies of their central banks. For example, in 2010, several European nations, most notably Greece, came close to defaulting on their government debt. This *eurozone crisis* reduced confidence and aggregate demand around the world, but the impact was much larger on Europe than on the United States. Thus, Europe and the United States not only have different monetary policies but also face different shocks. ■

The Taylor Principle

How much should the nominal interest rate set by the central bank respond to changes in inflation? The dynamic *AD–AS* model does not give a definitive answer, but it does offer an important guideline.

Recall the equation for monetary policy:

$$i_t = \pi_t + \rho + \theta_\pi (\pi_t - \pi_t^*) + \theta_Y \left(Y_t - \bar{Y}_t\right),$$

where θ_π and θ_Y are parameters that measure how much the interest rate set by the central bank responds to inflation and output. In particular, according to this equation, a 1-percentage-point increase in inflation π_t induces an increase in the nominal interest rate i_t of $1 + \theta_\pi$ percentage points. Because we assume that θ_π is greater than zero, whenever inflation increases, the central bank raises the nominal interest rate by an even larger amount.

The assumption that $\theta_\pi > 0$ has important implications for the behavior of the real interest rate. Recall that the real interest rate is $r_t = i_t - E_t\pi_{t+1}$. With our assumption of adaptive expectations, it can also be written as $r_t = i_t - \pi_t$. As a result, if an increase in inflation π_t leads to a greater increase in the nominal interest rate i_t, it leads to an increase in the real interest rate r_t as well. As you may recall from earlier in this chapter, this fact was a key part of our explanation for why the dynamic aggregate demand curve slopes downward.

Imagine, however, that the central bank behaved differently and, instead, increased the nominal interest rate by less than the increase in inflation. In this case, the monetary policy parameter θ_π would be less than zero. This change would profoundly alter the model. Recall that the dynamic aggregate demand equation is

$$Y_t = \bar{Y}_t - [\alpha\theta_\pi / (1 + \alpha\theta_Y)] (\pi_t - \pi_t^*) + [1/(1 + \alpha\theta_Y)] \varepsilon_t.$$

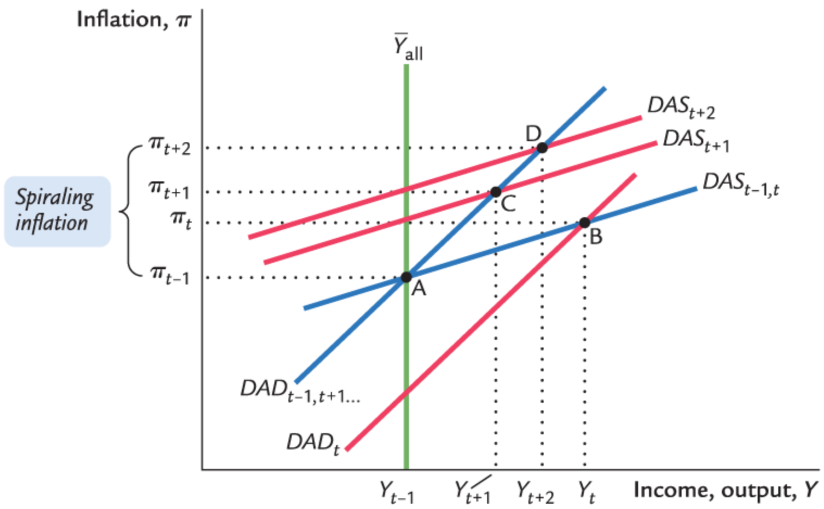
If θ_π is negative, then an increase in inflation increases the quantity of output demanded. To understand why, keep in mind what is happening to the real interest rate. If an increase in inflation leads to a smaller increase in the nominal interest rate (because $\theta_\pi < 0$), the real interest rate decreases. The lower real interest rate reduces the cost of borrowing, which in turn increases the quantity of goods and services demanded. Thus, a negative value of θ_π means the dynamic aggregate demand curve slopes upward.

An economy with $\theta_\pi < 0$ and an upward-sloping *DAD* curve can run into some serious problems. In particular, inflation can become unstable. Suppose, for example, there is a positive shock to aggregate demand that lasts for only a single period. Normally, such an event would have only a temporary effect on the economy, and the inflation rate would over time return to its target (similar to the analysis illustrated in Figure 15-9). If θπ<0,$\theta_\pi < 0$, however, events unfold very differently:

1. The positive demand shock increases output and inflation in the period in which it occurs.
2. Because expectations are determined adaptively, higher inflation increases expected inflation.
3. Because firms set their prices based in part on expected inflation, higher expected inflation leads to higher actual inflation in subsequent periods (even after the demand shock has dissipated).
4. Higher inflation causes the central bank to raise the nominal interest rate. But because θπ<0,$\theta_\pi < 0$, the central bank increases the nominal interest rate by less than the increase in inflation, so the real interest rate declines.
5. The lower real interest rate increases the quantity of goods and services demanded above the natural level of output.
6. With output above its natural level, firms face higher marginal costs, and inflation rises yet again.
7. The economy returns to step 2.

The economy finds itself in a vicious circle of ever-higher inflation and expected inflation. Inflation spirals out of control.

Figure 15-13 illustrates this process. Suppose in period *t* there is a one-time positive shock to aggregate demand. That is, for one period only, the dynamic aggregate demand curve shifts to the right, to DADt; DAD_t; in the next period, it returns to its original position. In period *t*, the economy moves from point A to point B. Output and inflation rise. In the next period, because higher inflation has increased expected inflation, the dynamic aggregate supply curve shifts upward, to DASt+1.DAS_{t+1}. The economy moves from point B to point C. But because the dynamic aggregate demand curve is now upward sloping, output remains above its natural level, even though demand shock has disappeared. Thus, inflation rises yet again, shifting the *DAS* curve farther upward in the next period, moving the economy to point D. And so on. Inflation continues to rise with no end in sight.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 15-13 The Importance of the Taylor Principle This figure shows the impact of a demand shock in an economy that does not satisfy the Taylor principle, so the dynamic aggregate demand curve is upward sloping. A demand shock moves the *DAD* curve to the right for one period, to DAD_t, and the economy moves from point A to point B. Both output and inflation increase. The rise in inflation increases expected inflation and, in the next period, shifts the dynamic aggregate supply curve upward to DAS_{t+1}. Therefore, in period $t+1$, the economy then moves from point B to point C. Because the *DAD* curve is upward sloping, output is still above the natural level, so inflation continues to increase. In period $t+2$, the economy moves to point D, where output and inflation are even higher. Inflation spirals out of control.

The dynamic *AD–AS* model leads to a strong conclusion: *for inflation to be stable, the central bank must respond to an increase in inflation with an even greater increase in the nominal interest rate.* This conclusion is sometimes called the **Taylor principle**, after economist John Taylor, who emphasized its importance in the design of monetary policy. (As we saw earlier, in his proposed Taylor rule, Taylor suggested that θ_π should equal 0.5.) Most of our analysis in this chapter assumed that the Taylor principle holds; that is, we assumed that $\theta_\pi > 0$. We can see now that there is good reason for a central bank to adhere to this guideline.

CASE STUDY

What Caused the Great Inflation?

In the 1970s, inflation in the United States got out of hand. As we saw in previous chapters, the inflation rate reached double-digit levels during that decade. Rising prices were considered the major economic problem of the time. In 1979, Paul Volcker, the recently appointed Fed chair, announced a change in monetary policy that eventually brought inflation back under control. Volcker and his successor, Alan Greenspan, then presided over low and stable inflation for the next quarter century.

The dynamic *AD–AS* model offers a new perspective on these events. According to research by monetary economists Richard Clarida, Jordi Galí, and Mark Gertler, the key is the Taylor principle. Clarida and colleagues examined the data on interest rates, output, and inflation and estimated the parameters of the monetary-policy rule. They found that the monetary policy of Volcker and Greenspan obeyed the Taylor principle, whereas earlier monetary policy did not. In particular, the parameter θ_π (which measures the responsiveness of interest rates to inflation in the monetary-policy rule) was estimated to be 0.72 during the Volcker–Greenspan regime after 1979, close to Taylor's proposed value of 0.5, but it was -0.14 during the pre-Volcker era from 1960 to 1978.[2] The negative value of θ_π during the pre-Volcker era means that monetary policy did not satisfy the Taylor principle. In other words, the pre-Volcker Fed was not responding strongly enough to inflation.

This finding suggests a cause of the great inflation of the 1970s. When the U.S. economy was hit by demand shocks (such as government spending on the Vietnam War) and supply shocks (such as the OPEC oil-price increases), the Fed raised the nominal interest rate in response to rising inflation but not by enough. Therefore, despite the increase in the nominal interest rate, the real interest rate fell. This insufficient monetary response failed to squash the inflation that arose from these shocks. Indeed, the decline in the real interest rate increased the quantity of goods and services demanded, thereby exacerbating the inflationary pressures. The problem of spiraling inflation was not solved until the monetary-policy rule was changed to include a more vigorous response of interest rates to inflation.

An open question is why policymakers were so passive in the earlier era. Here are some conjectures from Clarida, Galí, and Gertler:

> Why is it that during the pre-1979 period the Federal Reserve followed a rule that was clearly inferior? Another way to look at the issue is to ask why it is that the Fed maintained persistently low short-term real rates in the face of high or rising inflation. One possibility . . . is that the Fed thought the natural rate of unemployment at this time was much lower than it really was (or equivalently, that the output gap was much smaller). . . .

Another somewhat related possibility is that, at that time, neither the Fed nor the economics profession understood the dynamics of inflation very well. Indeed, it was not until the mid to late 1970s that intermediate textbooks began emphasizing the absence of a long-run trade-off between inflation and output. The ideas that expectations may matter in generating inflation and that credibility is important in policymaking were simply not well established during that era. What all this suggests is that in understanding historical economic behavior, it is important to take into account the state of policymakers' knowledge of the economy and how it may have evolved over time. ■

15-5 Conclusion: Toward DSGE Models

If you take more advanced courses in macroeconomics, you will likely learn about a class of models called dynamic, stochastic, general equilibrium models, often abbreviated as DSGE models. These models are *dynamic* because they trace the paths of variables over time. They are *stochastic* because they incorporate the inherent randomness of economic life. They are *general equilibrium* because they take into account that everything depends on everything else. In many ways, they are the state-of-the-art models for analyzing short-run economic fluctuations.

The dynamic *AD–AS* model in this chapter is a simplified version of these DSGE models. Unlike analysts using advanced DSGE models, we have not started with the optimizing decisions of households and firms that underlie macroeconomic relationships. But the macro relationships that this chapter has posited are similar to those found in more sophisticated DSGE models. The dynamic *AD–AS* model is a good stepping-stone between the basic model of aggregate demand and aggregate supply we saw in earlier chapters and the more complex DSGE models you might see in a more advanced course.[3]

The dynamic *AD–AS* model also yields some important lessons. It shows how various macroeconomic variables—output, inflation, and real and nominal interest rates—respond to shocks and interact with one another over time. It demonstrates that, in the design of monetary policy, central banks face a tradeoff between variability in inflation and variability in output. Finally, it suggests that central banks need to respond vigorously to inflation to prevent it from getting out of control. If you ever find yourself running a central bank, these are good lessons to keep in mind.

CHAPTER 16

Alternative Perspectives on Stabilization Policy

> The Federal Reserve's job is to take away the punch bowl just when the party gets going.
>
> —William McChesney Martin

> What we need is not a skilled monetary driver of the economic vehicle continuously turning the steering wheel to adjust to the unexpected irregularities of the route, but some means of keeping the monetary passenger who is in the back seat as ballast from occasionally leaning over and giving the steering wheel a jerk that threatens to send the car off the road.
>
> —Milton Friedman

How should government policymakers respond to the business cycle? The two quotations above—the first from a former Fed chair, the second from a prominent Fed critic—show the diversity of opinion over how this question is best answered.

Some economists, such as William McChesney Martin, view the economy as inherently unstable. They argue that the economy experiences frequent shocks to aggregate demand and aggregate supply. Unless policymakers use monetary and fiscal policy to stabilize the economy, these shocks will lead to unnecessary and inefficient fluctuations in output, unemployment, and inflation. According to the popular saying, macroeconomic policy should "lean against the wind," stimulating the economy when it is depressed and slowing the economy when it is overheated.

Other economists, such as Milton Friedman, view the economy as naturally stable. They blame bad economic policies for the large and inefficient fluctuations that sometimes occur. They argue that policy should not try to fine-tune the economy. Instead, policymakers should admit their limited abilities and be satisfied if they do no harm.

This debate has persisted for decades, with numerous protagonists advancing various arguments for their positions. The fundamental issue is how policymakers should use the theory of short-run fluctuations developed in the preceding chapters. In this chapter we ask two questions that arise in this debate. First, should monetary and fiscal policy actively try to stabilize the economy, or should policy remain passive? Second, should policymakers be free to use their discretion to respond to changing economic conditions, or should they be committed to a policy rule?

16-1 Should Policy Be Active or Passive?

Policymakers in the federal government view economic stabilization as one of their main responsibilities. Analyzing macroeconomic policy is a regular duty of the Fed, the Council of Economic Advisers, the Congressional Budget Office, and other government agencies. As we have seen in the preceding chapters, monetary and fiscal policy can exert a powerful impact on aggregate demand and, thereby, on inflation and unemployment. When Congress is considering a change in fiscal policy, or when the Fed is considering a change in monetary policy, foremost in the discussion are how the change will influence inflation and unemployment and whether aggregate demand needs to be stimulated or restrained.

Although the government has long conducted monetary and fiscal policy, the view that it should use these policy instruments to stabilize the economy is more recent. The Employment Act of 1946 was a landmark piece of legislation in which the government first held itself accountable for macroeconomic performance. The act states that "it is the continuing policy and responsibility of the Federal Government to . . . promote full employment and production." This law was written when the memory of the Great Depression was still fresh. The lawmakers who wrote it believed, as many economists do, that without an active government role in the economy, events like the Great Depression could occur regularly.

To many economists the case for active government policy is clear and simple. Recessions are periods of high unemployment, low incomes, and increased hardship. The model of aggregate demand and aggregate supply shows how shocks to the economy can cause recessions. It also shows how monetary and fiscal policy can respond to these shocks and prevent, or at least soften, recessions. These economists consider it wasteful not to use these policy instruments to stabilize the economy.

Other economists are critical of the government's attempts at stabilization. These critics argue that the government should take a hands-off approach to macroeconomic policy. At first, this view might seem surprising. If our model shows how to prevent or reduce the severity of recessions, why do these critics want the government to refrain from using monetary and fiscal policy to stabilize the economy? To find out, let's consider their arguments.

Lags in the Implementation and Effects of Policies

Economic stabilization would be easy if the effects of policy were immediate. Making policy would be like driving a car: policymakers would simply adjust their instruments to keep the economy on the desired path.

Making economic policy, however, is less like driving a car than it is like piloting a large ship. A car changes direction almost immediately after the steering wheel is turned. A ship changes course long after the pilot adjusts the rudder, and once the ship starts to turn, it continues turning long after the rudder is returned to normal. A novice pilot is likely to oversteer and, after noticing the mistake, overreact by steering too much in the opposite direction. The ship's path could become unstable, as the novice responds to previous mistakes by making larger and larger corrections.

Like a ship's pilot, economic policymakers face the problem of long lags. Indeed, the problem for policymakers is even more difficult because the lengths of the lags are hard to predict. The long and variable lags complicate the conduct of monetary and fiscal policy.

Economists distinguish between two lags that are relevant for the conduct of stabilization policy: the inside lag and the outside lag. The **inside lag** is the time between a shock to the economy and a policy action responding to that shock. This lag arises because it takes time for policymakers to recognize that a shock has occurred and to put appropriate policies into effect. The **outside lag** is the time between a policy action and its influence on the economy. This lag arises because policies do not immediately influence spending, income, and employment.

A long inside lag is a central problem with using fiscal policy for economic stabilization. This is especially true in the United States, where changes in spending or taxes require the approval of the president and both houses of Congress. The slow and cumbersome legislative process often leads to delays, making fiscal policy an imprecise tool for stabilizing the economy. This inside lag is shorter in countries with parliamentary systems, such as the United Kingdom, where the party in power can enact policy changes more rapidly.

Monetary policy has a shorter inside lag than fiscal policy because a central bank can decide on and implement a policy change in less than a day, but monetary policy has a larger outside lag. Monetary policy works by changing the money supply and interest rates, which influence investment and aggregate demand. Because many firms make investment plans far in advance, a change in monetary policy is thought not to affect economic activity until about six months after it is made.

The long and variable lags associated with monetary and fiscal policy certainly make stabilizing the economy more difficult. Advocates of passive policy argue that, because of these lags, successful stabilization policy is almost impossible. Indeed, attempts to stabilize the economy can be destabilizing. Suppose the economy's condition changes between the beginning of a policy action and its impact on the economy. In this case, active policy may end up stimulating the economy when it is heating up or depressing the economy when it is cooling off. Advocates of active policy admit that such lags require policymakers to be cautious. But, they argue, these lags do not mean that policy should be completely passive, especially in the face of a severe and protracted economic downturn.

Some policies, called **automatic stabilizers**, are designed to reduce the lags associated with stabilization policy. Automatic stabilizers are policies that stimulate or depress the economy when necessary without any deliberate policy change. For example, the system of income taxes automatically reduces taxes when the economy goes into a recession: without any change in the tax laws, individuals and corporations pay less tax when their incomes fall. Similarly, the unemployment-insurance and welfare systems automatically raise transfer payments when the economy moves into a recession because more people apply for benefits. One can view these automatic stabilizers as a type of fiscal policy without any inside lag.

The Difficult Job of Economic Forecasting

Because policy influences the economy only after a long lag, successful stabilization policy requires the ability to accurately predict future economic conditions. If we cannot predict whether the economy will be in a boom or a recession in six months or a year, we cannot evaluate whether monetary and fiscal policy should now be trying to expand or contract aggregate demand. Unfortunately, economic developments are often unpredictable.

Dana Fradon/The New Yorker/Conde Nast/
The Cartoon Bank

“It’s true, Caesar. Rome is declining, but I expect it to pick up in the next quarter.”

One way forecasters try to look ahead is with *leading indicators*. As we discussed in Chapter 10, a leading indicator is a data series that fluctuates in advance of the economy. A large fall in a leading indicator signals that a recession is more likely to occur in the coming months.

Another way forecasters look ahead is with *macroeconometric models*, which have been developed both by government agencies and by private firms. A macroeconometric model is a model that describes the economy quantitatively rather than just qualitatively. Many of these models are more complicated and realistic versions of the dynamic model of aggregate demand and aggregate supply in Chapter 15. The economists who build macroeconometric models use historical data to estimate a model’s parameters. Once a model is built, economists can simulate the effects of alternative policies. The model can also be used for forecasting. After the model’s user makes assumptions about the path of the exogenous variables, such as monetary policy, fiscal policy, and oil prices, the model yields predictions about unemployment, inflation, and other endogenous variables. Keep in mind, however, that the validity of these predictions is only as good as the model and the forecasters’ assumptions about the exogenous variables.

CASE STUDY

Mistakes in Forecasting

“Light showers, bright intervals, and moderate winds.” This was the forecast offered by the renowned British national weather service on October 14, 1987. The next day Britain was hit by its worst storm in over two centuries.

Like weather forecasts, economic forecasts are inputs to private and public decisionmaking. Business executives rely on forecasts when deciding how much to produce and how much to invest in plant and equipment. Government policymakers rely on them when developing economic policies. Unfortunately, like weather forecasts, economic forecasts are far from precise.

The most severe downturn in U.S. history, the Great Depression of the 1930s, caught forecasters completely by surprise. Even after the stock market crash of 1929, they remained confident that the economy would not suffer a substantial setback. In late 1931, when the economy was clearly in bad shape, the eminent economist Irving Fisher predicted that it would recover quickly. Subsequent events showed that these forecasts were much too optimistic: the unemployment rate continued to rise until 1933, when it hit 25 percent, and it remained elevated for the rest of the decade.[1]

Figure 16-1 shows how forecasters did during the recession of 2008–2009, the most severe economic downturn in the United States since the Great Depression. This figure shows the actual unemployment rate (in red) and several attempts to predict it for the following five quarters (in green). You can see that the forecasters did well when predicting unemployment one or two quarters ahead. The more distant forecasts, however, were often inaccurate. The November 2007 Survey of Professional Forecasters predicted a slowdown, but only a modest one: the U.S. unemployment rate was projected to increase from 4.7 percent in the fourth quarter of 2007 to 5.0 percent in the fourth quarter of 2008. By the May 2008 survey, the forecasters had raised their predictions for unemployment at the end of the year, but only to 5.5 percent. In fact, the unemployment rate was 6.9 percent in the last quarter of 2008. The forecasters became more pessimistic as the recession unfolded—but still not pessimistic enough. In November 2008, they predicted that the unemployment rate would rise to 7.7 percent in the fourth quarter of 2009. In fact, it rose to about 10 percent.

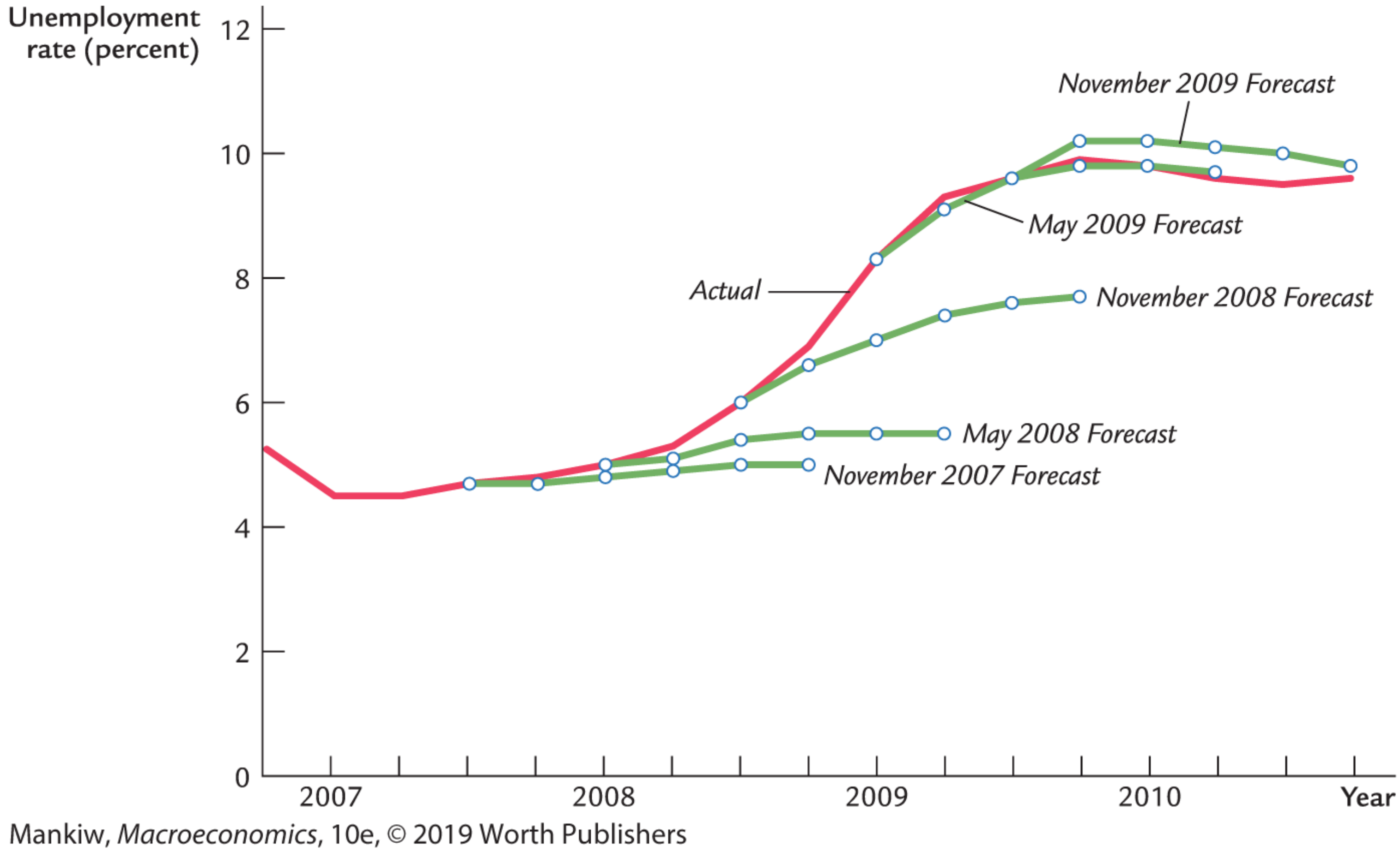


FIGURE 16-1 The Failure of Forecasting During the Great Recession The red line shows the actual unemployment rate from 2007 to 2010. The green lines show the unemployment rate predicted at various points in time. For each forecast, the symbols mark the current unemployment rate and the forecast for the subsequent five quarters. Note that the forecasters failed to predict the substantial rise in the unemployment rate.

Data from: The unemployment rate is from the U.S. Department of Labor. The predicted unemployment rate is the median forecast in the Survey of Professional Forecasters.

The Great Depression of the 1930s and the Great Recession of 2008–2009 show that the most dramatic economic events are often unpredictable. While private and public decisionmakers have little choice but to rely on economic forecasts, they must remember that these forecasts come with a large margin of error. ■

Ignorance, Expectations, and the Lucas Critique

The prominent economist Robert Lucas once wrote, “As an advice-giving profession we are in way over our heads.” Even many of those who advise policymakers would agree with this assessment. Economics is a young science, and there is still much that we do not know. Economists cannot be completely confident when assessing the effects of alternative policies. This ignorance suggests that economists should be cautious when offering policy advice.

In his writings on macroeconomic policymaking, Lucas has emphasized that economists need to pay more attention to the issue of how people form expectations of the future. Expectations play a key role in the economy because they influence all sorts of behavior. For instance, households decide how much to consume based on how much they expect to earn in the future, and firms decide how much to invest based on their expectations of future profitability. These expectations depend on many things, but one factor, according to Lucas, is especially important: the policies being pursued by the government. When policymakers estimate the effect of a policy change, therefore, they need to know how people's expectations will respond to the change. Lucas has argued that traditional methods of policy evaluation—such as those that rely on standard macroeconometric models—do not adequately take into account the impact of policy on expectations. This criticism of traditional policy evaluation is called the **Lucas critique**.[2]

An example of the Lucas critique arises in the analysis of disinflation. As you may recall from Chapter 14, the cost of reducing inflation is often measured by the sacrifice ratio, which is the number of percentage points of GDP that must be forgone to reduce inflation by 1 percentage point. Because estimates of the sacrifice ratio are often large, they have led some economists to argue that policymakers should learn to live with inflation rather than incur the large cost of reducing it.

According to advocates of the rational-expectations approach, however, these estimates of the sacrifice ratio are unreliable because they are subject to the Lucas critique. Traditional estimates of the sacrifice ratio are based on adaptive expectations—that is, on the assumption that expected inflation depends on past inflation. Adaptive expectations may be a reasonable premise in some circumstances, but if the policymakers make a credible change in policy, workers and firms setting wages and prices should respond rationally by adjusting their expectations of inflation appropriately. This change in inflation expectations would quickly alter the short-run tradeoff between inflation and unemployment. As a result, reducing inflation could be much less costly than traditional estimates of the sacrifice ratio suggest.

The Lucas critique leaves us with two lessons. The narrow lesson is that economists evaluating alternative policies need to consider how policy affects expectations and, thereby, behavior. The broad lesson is that policy evaluation is hard, so economists engaged in this task should show the requisite humility.

The Historical Record

In judging whether government policy should play an active or passive role in the economy, we must give some weight to the historical record. If the economy has experienced many large shocks to aggregate supply and aggregate demand, and if policy has successfully insulated the economy from these shocks, then the case for active policy should be clear. Conversely, if the economy has experienced few large shocks, and if the fluctuations we have observed can be traced to inept economic policy, then the case for passive policy should be clear. In other words, our view of stabilization policy should be influenced by whether policy has historically been stabilizing or destabilizing. For this reason, the debate over macroeconomic policy often turns into a debate over macroeconomic history.

Yet history does not settle the debate over stabilization policy. Disagreements over history arise because it is hard to identify the sources of economic fluctuations. The historical record often permits more than one interpretation.

The Great Depression is a case in point. Economists' views on macroeconomic policy are often related to their views on the cause of the Depression. Some economists believe that a large contractionary shock to private spending caused the Depression. They assert that policymakers should have responded by using the tools of monetary and fiscal policy to stimulate aggregate demand. Other economists believe that the large fall in the money supply caused the Depression. They assert that the Depression would have been avoided if the Fed had been pursuing a passive monetary policy of increasing the money supply at a steady rate. Hence, depending on one's beliefs about its cause, the Great Depression can be viewed either as an example of why active monetary and fiscal policy is necessary or as an example of why it is dangerous.

CASE STUDY

Is the Stabilization of the Economy a Figment of the Data?

Keynes wrote *The General Theory* in the 1930s, and in the wake of the Keynesian revolution, governments worldwide began to view stabilization as a primary responsibility. Some economists believe that the development of Keynesian theory has had a profound influence on the behavior of the economy. Comparing data from before World War I and after World War II, they find that real GDP and unemployment have become much more stable. This, some Keynesians claim, is the best argument for active stabilization policy: it has worked.

In a series of provocative and influential papers, economist Christina Romer has challenged this assessment of the historical record. She argues that the measured reduction in volatility reflects not an improvement in economic policy and performance but rather an improvement in the economic data. The older data are much less accurate than the newer data. Romer claims that the higher volatility of unemployment and real GDP reported for the period before World War I is largely a figment of the data.

Romer uses various techniques to make her case. One is constructing more accurate data for the earlier period. This task is difficult because data sources are not readily available. A second way is constructing *less* accurate data for the recent period—that is, data that are comparable to the older data and thus suffer from the same imperfections. After constructing new "bad" data, Romer finds that the recent period appears almost as volatile as the early period, suggesting that the volatility of the early period may be an artifact of how the data were assembled.

Romer's work is part of the debate over whether macroeconomic policy has improved the performance of the economy. Although her work remains controversial, most economists now believe that the economy in the aftermath of the Keynesian revolution was only slightly more stable than it had been before.[3] ■

CASE STUDY

How Does Policy Uncertainty Affect the Economy?

When monetary and fiscal policymakers actively try to control the economy, the future course of economic policy is often uncertain. Policymakers do not always make their intentions clear. Moreover, because the policy outcome can be the result of a divisive, contentious, and unpredictable political process, the public has every reason to be unsure about what policy decisions will end up being made.

In recent research, economists Scott Baker, Nicholas Bloom, and Steve Davis studied the effects of policy uncertainty. Baker, Bloom, and Davis began by constructing an index that measures how the amount of policy uncertainty changes over time. Their index has three components.

The first component is derived from reading newspaper articles. Starting in January 1985, they examined ten major papers for terms related to economic and policy uncertainty. In particular, they searched for articles containing the term *uncertainty* or *uncertain*, the term *economic* or *economy*, and at least one of the following terms: *congress*, *legislation*, *white house*, *regulation*, *federal reserve*, or *deficit*. The more articles there were that included terms in all three categories, the higher the index of policy uncertainty.

The second component of the index is based on the number of temporary provisions in the federal tax code. Baker, Bloom, and Davis reasoned that “temporary tax measures are a source of uncertainty for businesses and households because Congress often extends them at the last minute, undermining stability in and certainty about the tax code.” The more temporary tax provisions there are and the larger the dollar magnitudes involved in the provisions, the higher the index of policy uncertainty.

The third component of the index is based on the amount of disagreement among private forecasters about several key variables related to macroeconomic policy. Baker, Bloom, and Davis assumed that the more private forecasters disagree about the future price level and future levels of government spending, the more uncertainty there is about monetary and fiscal policy. That is, the greater the dispersion in these private forecasts, the higher the level of the policy uncertainty index.

Figure 16-2 shows the index derived from these three components. The index spikes upward, indicating an increase in policy uncertainty, when there is a significant foreign policy event (such as a war or terrorist attack), when there is an economic crisis (such as the Black Monday stock market crash or the bankruptcy of the large investment bank Lehman Brothers), or when there is a major political event (such as the election of a new president).

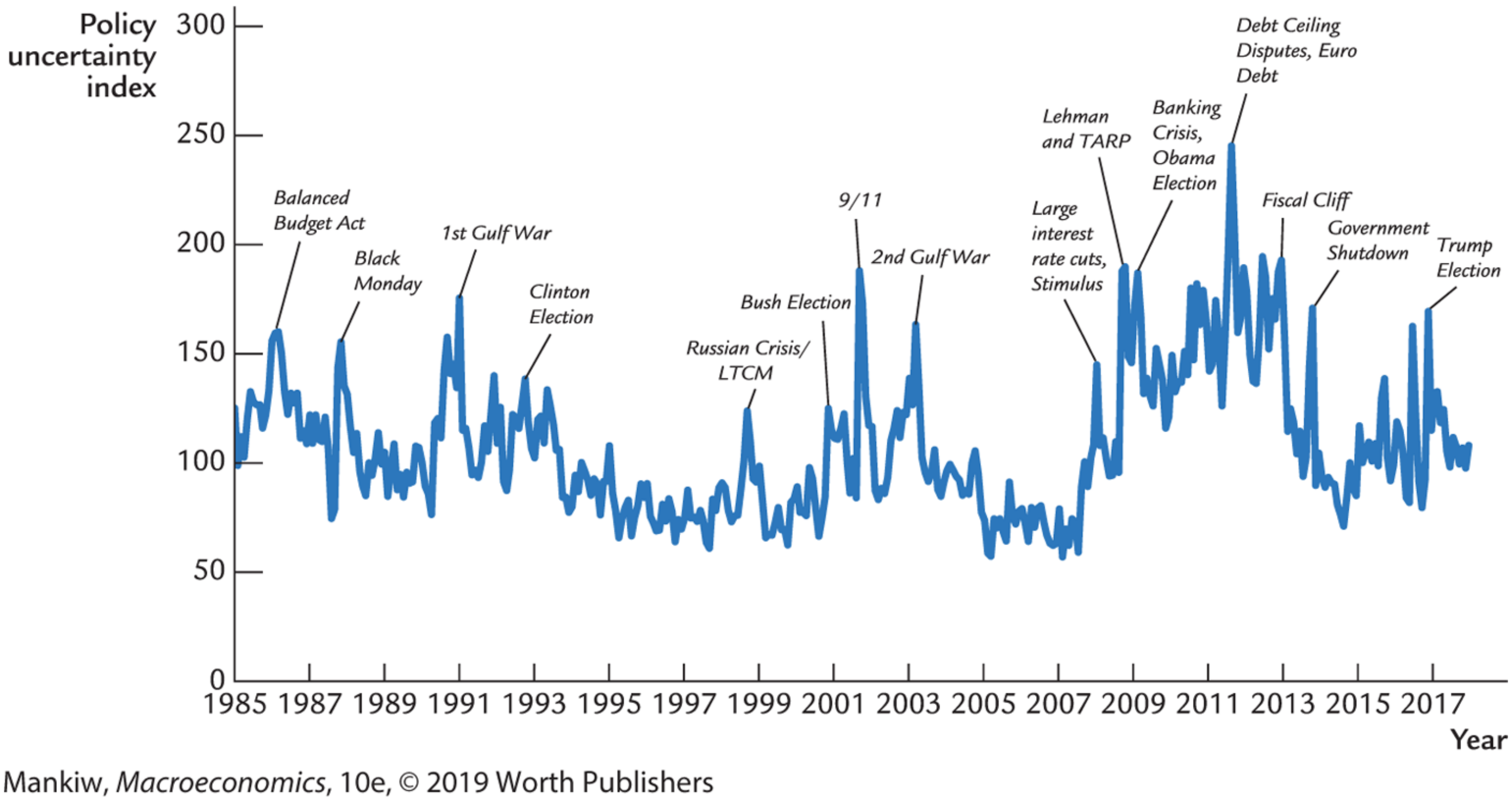


FIGURE 16-2 An Index of Policy Uncertainty Various kinds of events cause uncertainty about policy to increase. Spikes in policy uncertainty may depress economic activity.

Data from: http://www.policyuncertainty.com/us_monthly.html.

With this index in hand, Baker, Bloom, and Davis then investigated how policy uncertainty correlates with economic performance. They found that higher uncertainty about economic policy depresses the economy. In particular, when economic policy uncertainty rises, investment, production, and employment are likely to decline over the next year (relative to their normal growth).

One possible explanation for this effect is that uncertainty may depress the aggregate demand for goods and services. When policy uncertainty increases, households and firms may delay large purchases until the uncertainty is resolved. For example, if a firm is considering building a new factory, and the profitability of the investment depends on what policy is pursued, the firm may wait until a policy decision is made. Such a delay is rational for the firm, but it contributes to a decline in aggregate demand, reducing the economy's output and raising unemployment.

To be sure, some policy uncertainty is inevitable. But policymakers should keep in mind that the amount of uncertainty is, to some degree, under their control and that heightened uncertainty appears to have adverse economic effects.[4] ■

16-2 Should Policy Be Conducted by Rule or Discretion?

A second topic debated among economists is whether economic policy should be conducted by rule or discretion. Policy is conducted by rule if policymakers announce in advance how policy will respond to various situations and commit themselves to following through on this announcement. Policy is conducted by discretion if policymakers are free to size up events as they occur and choose whatever policy they deem appropriate at the time.

The debate over rules versus discretion is distinct from the debate over passive versus active policy. Policy can be conducted by rule and yet be either passive or active. For example, a passive policy rule might specify steady growth in the money supply of 3 percent per year. An active policy rule might specify

Money Growth=3%+(Unemployment Rate−6%).

$$\text{Money Growth} = 3\% + (\text{Unemployment Rate} - 6\%).$$

Under this rule, the money supply grows at 3 percent if the unemployment rate is 6 percent, but for every percentage point by which the unemployment rate exceeds 6 percent, money growth increases by an extra percentage point. This rule tries to stabilize the economy by raising money growth during recessions.

We begin this section by discussing why policy might be improved by a commitment to a policy rule. We then examine several possible policy rules.

Distrust of Policymakers and the Political Process

Some economists believe that economic policy is too important to be left to the discretion of policymakers. Although this view is more political than economic, evaluating it is central to how we judge the role of economic policy. If politicians are incompetent or opportunistic, then we may not want to give them the discretion to use the powerful tools of monetary and fiscal policy.

Incompetence in economic policy arises for several reasons. Some economists view the political process as erratic, perhaps because it reflects the shifting power of special-interest groups. In addition, macroeconomics is complicated, and politicians often do not have sufficient knowledge of it to make informed judgments. This ignorance allows charlatans to propose incorrect but superficially appealing solutions to complex problems. The political process often cannot weed out the advice of charlatans from that of competent economists.

Opportunism in economic policy arises when the objectives of policymakers conflict with the well-being of the public. Some economists fear that politicians use macroeconomic policy to further their own electoral ends. If citizens vote on the basis of economic conditions prevailing at the time of the election, then politicians have an incentive to pursue policies that will make the economy look good during election years. A president might cause a recession soon after coming into office to lower inflation and then stimulate the economy as the next election approaches to lower unemployment; this would ensure that both inflation and unemployment are low on Election Day. Manipulation of the economy for electoral gain, called the **political business cycle**, has been the subject of extensive research by economists and political scientists.[5]

Distrust of the political process leads some economists to advocate placing economic policy outside the realm of politics. Some have proposed constitutional amendments, like a balanced-budget amendment, that would tie the hands of legislators and insulate the economy from both incompetence and opportunism. We discuss some potential problems with a balanced-budget amendment in the next chapter.

The Time Inconsistency of Discretionary Policy

If we assume that we can trust our policymakers, discretion at first glance appears superior to a policy rule. Discretionary policy is flexible. As long as policymakers are intelligent and benevolent, there might appear to be little reason to deny them flexibility in responding to changing conditions.

Yet a case for rules over discretion arises from the problem of **time inconsistency** of policy. In some situations, policymakers may want to announce in advance the policy they will follow to influence the expectations of private decisionmakers. But later, after the private decisionmakers have acted on the basis of their expectations, the policymakers may be tempted to renege on their announcement. Understanding that policymakers may be inconsistent over time, private decisionmakers distrust policy announcements. In this situation, policymakers may want to commit themselves to a policy rule to make their announcements credible.

Time inconsistency is illustrated most simply with a political rather than an economic example—specifically, public policy about negotiating with terrorists over the release of hostages. The announced policy of many nations is that they will not negotiate over hostages. This announcement is intended to deter terrorists: if there is nothing to be gained from kidnapping hostages, rational terrorists won't kidnap any. In other words, the purpose of the announcement is to influence the expectations of terrorists and, thereby, their behavior.

But, in fact, unless the policymakers are credibly committed to the policy, the announcement has little effect. Terrorists know that once hostages are taken, policymakers face an overwhelming temptation to make some concession to obtain the hostages' release. The only way to deter rational terrorists is to take away the discretion of policymakers and commit them to a rule of never negotiating. If policymakers were truly unable to make concessions, the incentive for terrorists to take hostages would be largely eliminated.

The same problem arises less dramatically in the conduct of monetary policy. Consider the dilemma of a central bank that cares about both inflation and unemployment. According to the Phillips curve, the tradeoff between inflation and unemployment depends on expected inflation. The central bank would prefer everyone to expect low inflation so that it will face a favorable tradeoff. To reduce expected inflation, the central bank might announce that low inflation is its paramount goal.

But an announcement of a policy of low inflation is by itself not credible. Once households and firms have formed their expectations of inflation and set wages and prices accordingly, the central bank has an incentive to renege on its announcement and pursue expansionary monetary policy to reduce unemployment. People understand the central bank's incentive to renege and may not believe the announcement in the first place. Just as a president facing a hostage crisis is tempted to negotiate their release, a central bank with discretion is tempted to inflate in order to reduce unemployment. And just as terrorists discount announced policies of never negotiating, households and firms discount announced policies of low inflation.

The surprising outcome of this analysis is that policymakers can sometimes better achieve their goals by removing their discretion. In the case of rational terrorists, fewer hostages will be taken and killed if policymakers are committed to the seemingly harsh rule of refusing to negotiate for hostages' freedom. In the case of monetary policy, there will be lower inflation without higher unemployment if the central bank is committed to a policy of zero inflation. (This conclusion about monetary policy is modeled more formally in the appendix to this chapter.)

The time inconsistency of policy arises in many other contexts:

- To encourage investment, the government announces that it will not tax income from capital. But after factories have been built, the government is tempted to renege on its promise to raise more tax revenue.
- To encourage research, the government announces that it will give a temporary monopoly to companies that discover new drugs. But after a drug has been discovered, the government is tempted to revoke the patent or to regulate the price to make the drug more affordable.
- To encourage good behavior, a parent announces that he will punish a child whenever the child breaks a rule. But after the child has misbehaved, the parent is tempted to forgive the transgression because punishment is unpleasant for both the parent and the child.
- To encourage you to work hard, your professor announces that this course will end with an exam. But

> after you have studied and learned the material, the professor is tempted to cancel the exam so that he won't have to grade it.

In each case, rational agents understand the incentive for the policymaker to renege, and this expectation affects their behavior. The solution is to take away the policymaker's discretion with a credible commitment to a policy rule.

CASE STUDY

Alexander Hamilton Versus Time Inconsistency

Time inconsistency has long been a problem associated with discretionary policy. In fact, it was one of the first problems that confronted Alexander Hamilton when President George Washington appointed him the first U.S. Secretary of the Treasury in 1789.

Hamilton faced the question of how to deal with the debts that the new nation had accumulated as it fought for its independence from Britain. When the revolutionary government incurred the debts, it promised to honor them when the war was over. But after the war, many Americans advocated defaulting on the debt because repaying the creditors would require taxation, which is always costly and unpopular.

Hamilton opposed the time-inconsistent policy of repudiating the debt. He realized that the nation would need to borrow again sometime in the future. In his *First Report on the Public Credit*, which he presented to Congress in 1790, he wrote:

> If the maintenance of public credit, then, be truly so important, the next inquiry which suggests itself is: By what means is it to be effected? The ready answer to which question is, by good faith; by a punctual performance of contracts. States, like individuals, who observe their engagements are respected and trusted, while the reverse is the fate of those who pursue an opposite conduct.

Thus, Hamilton proposed that the nation make a commitment to the policy rule of honoring its debts.

The policy rule that Hamilton proposed has continued for over two centuries. Today, unlike in Hamilton's time, when Congress debates spending priorities, no one seriously proposes defaulting on the public debt as a way to reduce taxes. In the case of public debt, Americans now agree that the government should be committed to a policy rule. ■

Rules for Monetary Policy

Even if we are convinced that policy rules are superior to discretion, the debate over macroeconomic policy is not over. If the Fed were to commit to a rule for monetary policy, what rule should it choose? Let's discuss three policy rules that various economists advocate.

Some economists, called **monetarists**, advocate that the Fed keep the money supply growing at a steady rate. The quotation at the beginning of this chapter from Milton Friedman—the most famous monetarist—

exemplifies this view of monetary policy. Monetarists believe that fluctuations in the money supply are responsible for most large fluctuations in the economy. They argue that slow and steady growth in the money supply would yield stable output, employment, and prices.

A monetarist rule might have prevented many of the fluctuations we have experienced historically, but most economists believe that it is not the best possible policy rule. Steady growth in the money supply stabilizes aggregate demand only if the velocity of money is stable. But sometimes the economy experiences shocks, such as shifts in money demand, that cause velocity to be unstable. Most economists believe that a policy rule needs to allow the money supply to adjust to various shocks to the economy.

A second policy rule that many economists advocate is nominal GDP targeting. Under this rule, the Fed announces a planned path for nominal GDP. If nominal GDP rises above the target, the Fed adjusts monetary policy to dampen aggregate demand. If it falls below the target, the Fed adjusts monetary policy to stimulate aggregate demand. Because a nominal GDP target allows monetary policy to adjust to changes in the velocity of money, most economists believe it would lead to greater stability in output and prices than a monetarist rule.

A third policy rule that is often advocated is **inflation targeting**. Under this rule, the Fed would announce a target for the inflation rate (usually a low one) and then adjust monetary policy when the actual inflation rate deviates from the target. Like nominal GDP targeting, inflation targeting insulates the economy from changes in the velocity of money. In addition, an inflation target has the political advantage of being easy to explain to the public.

Notice that all these rules are expressed in terms of some nominal variable—the money supply, nominal GDP, or the price level. One can also imagine policy rules expressed in terms of real variables. For example, the Fed might try to target an unemployment rate of 5 percent. The problem with such a rule is that no one knows exactly what the natural rate of unemployment is. If the Fed chose a target for unemployment below the natural rate, the result would be accelerating inflation. Conversely, if the Fed chose a target for unemployment above the natural rate, the result would be accelerating deflation. For this reason, economists rarely advocate rules for monetary policy expressed solely in terms of real variables, even though real variables such as unemployment and real GDP are the best measures of economic performance.

CASE STUDY

Inflation Targeting: Rule or Constrained Discretion?

Beginning in the late 1980s, many of the world's central banks—including those of Australia, Canada, Finland, Israel, New Zealand, Sweden, and the United Kingdom—started to adopt some form of inflation targeting. Sometimes inflation targeting takes the form of a central bank announcing its policy intentions. At other times it takes the form of a national law that spells out the goals of monetary policy. For example, the Reserve Bank of New Zealand Act of 1989 told the central bank "to formulate and implement monetary policy directed to the economic objective of achieving and maintaining stability in the general level of prices." The act omitted any mention of other competing objectives, such as stability in output, employment, interest rates, or exchange rates.

Should we interpret inflation targeting as a type of commitment to a policy rule? Not completely. In countries that have adopted inflation targeting, central banks are left with some discretion. Inflation targets are usually set as a range—an inflation rate of 1 to 3 percent, for instance—rather than a specific number. The central bank can choose where in the range it wants to be: it can stimulate the economy and be near the top of the range or dampen the economy and be near the bottom. In addition, the central bank is sometimes allowed to adjust its target for inflation, at least temporarily, if an exogenous event (such as an easily identified supply shock) pushes inflation outside the announced range.

In light of this flexibility, what is the purpose of inflation targeting? Although inflation targeting leaves the central bank with some discretion, the policy constrains how this discretion is used. When a central bank is told simply to "do the right thing," it is hard to hold the central bank accountable because people can argue forever about what the right thing is in any particular circumstance. By contrast, when a central bank has announced an inflation target, the public can more easily judge whether the central bank is meeting its objectives. Thus, although inflation targeting does not tie the hands of the central bank, it increases the transparency of monetary policy and thereby makes central bankers more accountable.[6]

The Fed was slow to adopt a policy of inflation targeting, but in 2012 it set for itself an inflation target of 2 percent. On its website, the Fed offers this explanation:

> The Federal Open Market Committee (FOMC) judges that inflation at the rate of 2 percent (as measured by the annual change in the price index for personal consumption expenditures, or PCE) is most consistent over the longer run with the Federal Reserve's mandate for price stability and maximum employment. Over time, a higher inflation rate would reduce the public's ability to make accurate longer-term economic and financial decisions. On the other hand, a lower inflation rate would be associated with an elevated probability of falling into deflation, which means prices and perhaps wages, on average, are falling—a phenomenon associated with very weak economic conditions. Having at least a small level of inflation makes it less likely that the economy will experience harmful deflation if economic conditions weaken. The FOMC implements monetary policy to help maintain an inflation rate of 2 percent over the medium term.

More recently, a debate has arisen as to whether 2 percent is the right target for inflation. For six years after the Great Recession of 2008–2009, the Fed kept the federal funds rate at zero, its lower bound. (The zero lower bound was discussed in Chapter 12.) Even in early 2018, as this book was going to press, this interest rate was only 1.4 percent. Some economists worry that, if the economy were to experience a contractionary shock, the Fed would not have much room to cut interest rates and stimulate aggregate demand. They argue that if the Fed had a higher inflation target—say, 4 percent—the normal level of interest rates would be higher (via the Fisher effect), and the Fed would have more ammunition to combat downturns when necessary. Defenders of the current policy argue that the Fed would suffer too great a loss in credibility if it switched to a 4 percent inflation target after convincing the public of its commitment to a 2 percent target. At least so far, the Fed has shown no interest in revising its target. ■

CASE STUDY

Central-Bank Independence

Suppose you were writing the constitution and laws for a country. Would you give the country’s president authority over the policies of the central bank? Or would you allow the central bank to make decisions free from political influence? In other words, assuming that monetary policy is made by discretion rather than by rule, who should exercise that discretion?

Countries differ in how they answer this question. In some countries, the central bank is a branch of the government; in others, the central bank is largely independent. In the United States, Fed governors are appointed by the president for 14-year terms and cannot be recalled if the president is unhappy with their decisions. This institutional structure gives the Fed a degree of independence similar to that of the U.S. Supreme Court.

Many researchers have investigated the effects of constitutional design on monetary policy. They have examined the laws of different countries to construct an index of central-bank independence. This index is based on various characteristics, such as the length of bankers’ terms, the role of government officials on the bank board, and the frequency of contact between the government and the central bank. The researchers then examined the correlation between central-bank independence and macroeconomic performance.

The results of these studies are striking: more independent central banks are strongly associated with lower and more stable inflation. Figure 16-3 shows a scatterplot of central-bank independence and average inflation for the period 1955 to 1988. Countries that had an independent central bank, such as Germany, Switzerland, and the United States, tended to have low average inflation. Countries that had central banks with less independence, such as New Zealand and Spain, tended to have higher average inflation.

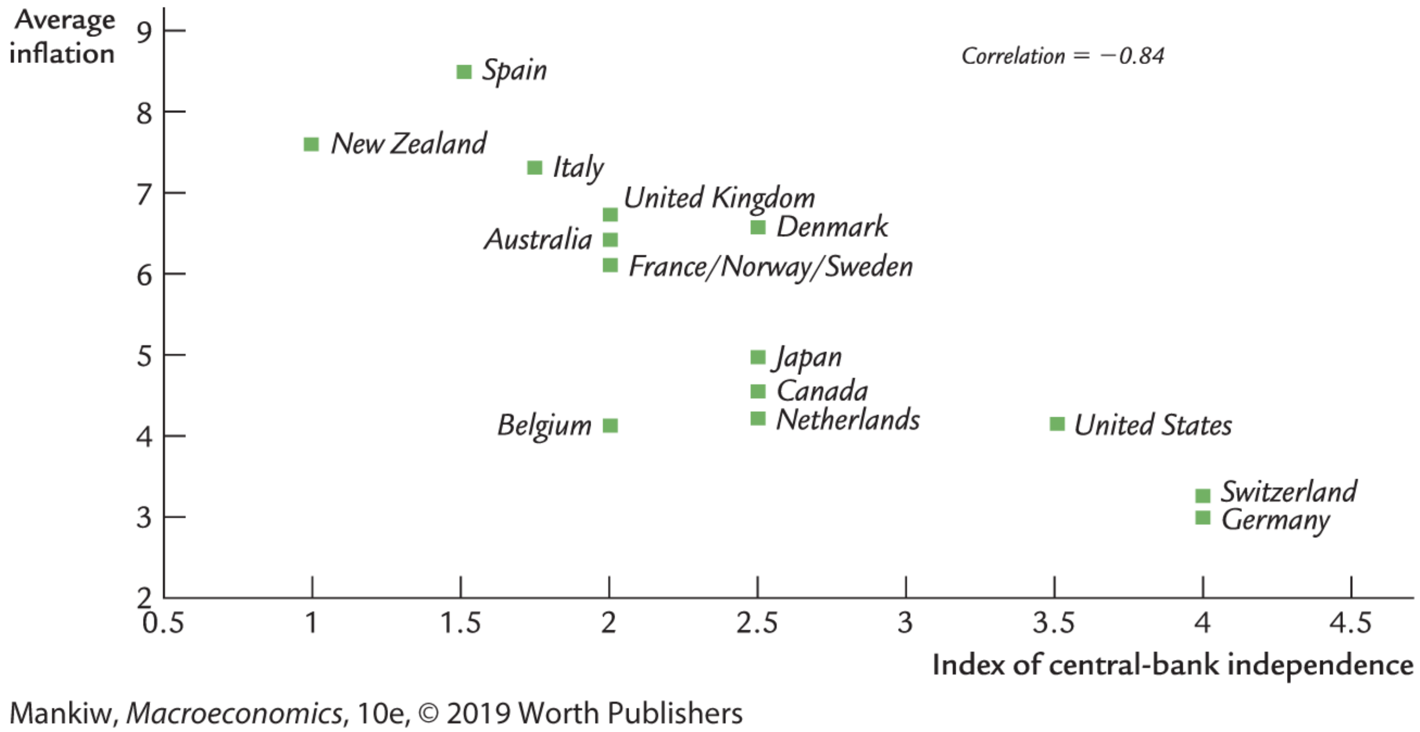


FIGURE 16-3 Inflation and Central-Bank Independence This scatterplot presents the international experience with central-bank independence. The evidence shows that more independent central banks tend to produce lower rates of inflation.

Data from: Figure 1a, page 155, of Alberto Alesina and Lawrence H. Summers, “Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence,” *Journal of Money, Credit, and Banking* 25 (May 1993): 151–162. Average inflation is for the period 1955–1988.

Researchers have also found that there is no relationship between central-bank independence and real economic activity. Central-bank independence is not correlated with average unemployment, the volatility of unemployment, the average growth of real GDP, or the volatility of real GDP. Central-bank independence appears to offer countries a free lunch: it has the benefit of lower inflation without any apparent cost. This finding has led some countries, such as New Zealand, to rewrite their laws to give their central banks greater independence.[7] ■

16-3 Conclusion

In this chapter we examined whether policy should take an active or a passive role in responding to economic fluctuations and whether policy should be conducted by rule or by discretion. There are many arguments on both sides of these questions. Perhaps the only clear conclusion is that there is no simple and compelling case for any particular view of macroeconomic policy. In the end, you must weigh the arguments, both economic and political, and decide for yourself what role the government should play in trying to stabilize the economy.

APPENDIX

Time Inconsistency and the Tradeoff Between Inflation and Unemployment

In this appendix, we examine more formally the time-inconsistency argument for rules rather than discretion. This analysis is relegated to an appendix because it requires some calculus.[8]

Suppose the Phillips curve describes the relationship between inflation and unemployment. Letting u denote the unemployment rate, u^n the natural rate of unemployment, π the rate of inflation, and $E\pi$ the expected rate of inflation, unemployment is determined by

$$u = u^n - \alpha(\pi - E\pi).$$

Unemployment is low when inflation exceeds expected inflation and high when inflation falls below expected inflation. The parameter α determines how much unemployment responds to surprise inflation.

Suppose also the central bank chooses the rate of inflation. In reality, central banks control inflation imperfectly using the tools of monetary policy. But it is a useful simplification to assume that the central bank controls inflation perfectly.

The central bank likes low unemployment and stable prices. We represent the cost of unemployment and inflation as

$$L(u,\ \pi) = u + \gamma\pi^2,$$

where the parameter γ represents how much the central bank dislikes inflation relative to unemployment. $L(u,\ \pi)$ is called the *loss function*. The central bank objective is to minimize the loss.

Having specified how the economy works and the central bank's objective, let's compare monetary policy made under a rule and under discretion.

We begin by considering policy under a rule. A rule commits the central bank to a particular level of inflation. As long as private agents understand that the central bank is committed to this rule, the expected level of inflation will be the level the central bank is committed to produce. Because expected inflation equals actual inflation $(E\pi = \pi)$, unemployment will be at its natural rate $(u = u^n)$.

What is the optimal rule? Because unemployment is at its natural rate regardless of the level of inflation legislated by the rule, there is no benefit to inflation. Therefore, the optimal rule requires that the central bank produce zero inflation.

Now let's consider discretionary monetary policy. Under discretion, the economy works as follows:

1. Private agents form their expectations of inflation $E\pi$.
2. The central bank chooses the actual level of inflation π.
3. Based on expected and actual inflation, unemployment is determined.

Under this arrangement, the central bank minimizes its loss $L(u,\ \pi)$ subject to the constraint that the Phillips curve imposes. When making its decision about the rate of inflation, the central bank takes expected inflation as already determined.

To find what outcome we would obtain under discretionary policy, we must examine what level of inflation the central bank would choose. By substituting the Phillips curve into the central bank's loss function, we obtain

$$L(u,\ \pi) = u^n - \alpha(\pi - E\pi) + \gamma\pi^2.$$

Notice that the central bank's loss is negatively related to unexpected inflation (the second term in the equation) and positively related to actual inflation (the third term). To find the level of inflation that minimizes this loss, differentiate with respect to π to obtain

$$dL / d\pi = -\alpha + 2\gamma\pi.$$

The loss is minimized when this derivative equals zero.[9] Solving for π, we get

$$\pi = \alpha / (2\gamma).$$

Whatever level of inflation private agents expected, this is the optimal level of inflation for the central bank to choose. Rational private agents understand the objective of the central bank and the constraint that the Phillips curve imposes. Thus, they expect that the central bank will choose this level of inflation. Expected inflation equals actual inflation $[E\pi=\pi=\alpha/(2\gamma)]$,$[E\pi = \pi = \alpha / (2\gamma)]$, and unemployment equals its natural rate (u=un).$(u = u^n)$.

Now compare the outcome under optimal discretion to the outcome under the optimal rule. In both cases, unemployment is at its natural rate. Yet discretionary policy produces more inflation than does policy under the rule. *Hence, optimal discretion is worse than the optimal rule*. This is true even though the central bank under discretion was attempting to minimize its loss, $L(u, \pi)$.$L(u,\ \pi)$.

It may seem strange that the central bank achieves a better outcome by being committed to a rule. Why can’t the central bank with discretion mimic the central bank committed to a zero-inflation rule? The answer is that the central bank is playing a game against private decisionmakers with rational expectations. Unless it is committed to a rule of zero inflation, the central bank cannot get private agents to expect zero inflation.

Suppose, for example, the central bank simply announces that it will follow a zero-inflation policy. The announcement won’t be credible. After private agents have formed their expectations of inflation, the central bank has the incentive to renege on its announcement in order to decrease unemployment. As we have just seen, once expectations are determined, the central bank’s optimal policy is to set inflation at $\pi=\alpha/(2\gamma)$, $\pi = \alpha / (2\gamma)$, regardless of $E\pi$.$E\pi$. Private agents understand the incentive to renege and therefore do not believe the announcement in the first place.

This theory of monetary policy has an important corollary. In one circumstance, the central bank with discretion achieves the same outcome as the central bank committed to a rule of zero inflation. If the central bank dislikes inflation much more than it dislikes unemployment (so that $\gamma$$\gamma$ is very large), inflation under discretion is near zero because the central bank has little incentive to inflate. This finding provides some guidance to those who appoint central bankers. An alternative to imposing a rule is to appoint an individual with a fervent distaste for inflation. Perhaps this is why even liberal politicians (Jimmy Carter, Bill Clinton) who are more concerned about unemployment than inflation sometimes appoint conservative central bankers (Paul Volcker, Alan Greenspan) who are more concerned about inflation.[10]

MORE PROBLEMS AND APPLICATIONS

1. In the 1970s in the United States, the inflation rate and the natural rate of unemployment both rose. Let’s use this model of time inconsistency to examine this phenomenon. Assume that policy is discretionary.

    a. In the model as developed so far, what happens to the inflation rate when the natural rate of unemployment rises?

    b. Let’s now change the model slightly by supposing that the Fed’s loss function is quadratic in both inflation and unemployment. That is,

$$L(u,\ \pi) = u^2 + \gamma\pi^2.$$

Follow steps similar to those in the text to solve for the inflation rate under discretionary policy.

c. Now what happens to the inflation rate when the natural rate of unemployment rises?

d. In 1979, President Jimmy Carter appointed the central banker Paul Volcker to head the Fed. Volcker had a strong distaste for inflation. According to this model, what should have happened to inflation and unemployment? Compare the model’s prediction to what actually happened.

CHAPTER 17

Government Debt and Budget Deficits

> Blessed are the young, for they shall inherit the national debt.
>
> —Herbert Hoover

> I think we ought to just go ahead and make "zillion" a real number. "Gazillion," too. A zillion could be ten million trillions, and a gazillion could be a trillion zillions. It seems to me it's time to do this.
>
> —George Carlin

When a government spends more than it collects in taxes, it has a budget deficit, which it finances by borrowing from the private sector or from foreign governments. The accumulation of past borrowing is the government debt.

Debate about the appropriate amount of government debt in the United States is as old as the country itself. Alexander Hamilton believed that "a national debt, if it is not excessive, will be to us a national blessing," while James Madison argued that "a public debt is a public curse." Indeed, the location of the nation's capital was chosen as part of a deal in which the federal government assumed the Revolutionary War debts of the states: because the northern states had larger outstanding debts, the capital was located in the South.

This chapter considers various aspects of the debate over government debt. We begin by looking at the numbers. Section 17-1 examines the size of the U.S. government debt and compares it with the historical and international record. It also takes a brief look at what the future may hold. Section 17-2 discusses why measuring changes in government indebtedness is not as straightforward as it might seem.

We then examine how government debt affects the economy. Section 17-3 describes the traditional view of government debt, according to which government borrowing reduces national saving and crowds out capital accumulation. This view is held by most economists and has been implicit in the discussion of fiscal policy throughout this book. Section 17-4 discusses an alternative view, called *Ricardian equivalence*. According to the Ricardian view, government debt does not influence national saving and capital accumulation. As we will see, the debate between the traditional and Ricardian views of government debt arises from disagreements over how consumers respond to the government's debt policy.

Section 17-5 then looks at other facets of the debate over government debt. It begins by discussing whether the government should always try to balance its budget and, if not, when a budget deficit or surplus is desirable. It also examines the effects of government debt on monetary policy, the political process, and a nation's role in the world economy.

This chapter provides the foundation for understanding government debt and budget deficits, but the story will not be completed until the next chapter. There we will examine the financial system more broadly, including the causes of financial crises. As we will see, excessive government debt can be at the center of such crises—a lesson that several European nations have recently learned, all too painfully.

17-1 The Size of the Government Debt

Let's begin by putting the government debt in perspective. In 2016, the debt of the U.S. federal government was $14.2 trillion. If we divide this number by 323 million, the population of the United States, we find that each person's share of the government debt was about $44,000. Obviously, this is not a trivial number; few people sneeze at $44,000. Yet compared with the roughly $2 million a typical person will earn over her working life, the government debt does not look like the catastrophe it is sometimes made out to be.

One way to judge the size of a government's debt is to compare it with the debt of other countries. Table 17-1 shows government debt for several major countries expressed as a percentage of each country's GDP. The figure here is net debt: the government's financial obligations less any financial assets that it holds. At the top of the list are the heavily indebted countries of Greece, Italy, and Japan, which have accumulated debt that exceeds annual GDP. At the bottom are Switzerland and Australia, which have accumulated little debt. The United States is near the middle of the pack. By international standards, the U.S. government is neither especially profligate nor especially frugal.

TABLE 17-1 How Indebted Are the World's Governments?

Country	Government Debt as a Percentage of GDP
Greece	149.1
Italy	129.9
Japan	125.5
Portugal	104.7
Belgium	96.5
United Kingdom	91.8
Spain	83.0
United States	81.3
France	79.0
Netherlands	40.6
Germany	39.9
Canada	31.0

Switzerland	5.5
Australia	−15.3
Data from: OECD Economic Outlook. Data are net financial liabilities as a percentage of GDP for 2016.	

Over the course of U.S. history, the indebtedness of the federal government has varied substantially. Figure 17-1 shows the ratio of the federal debt to GDP since 1791. The government debt, relative to the size of the economy, has varied from close to zero in the 1830s to a maximum of 106 percent of GDP in 1946.

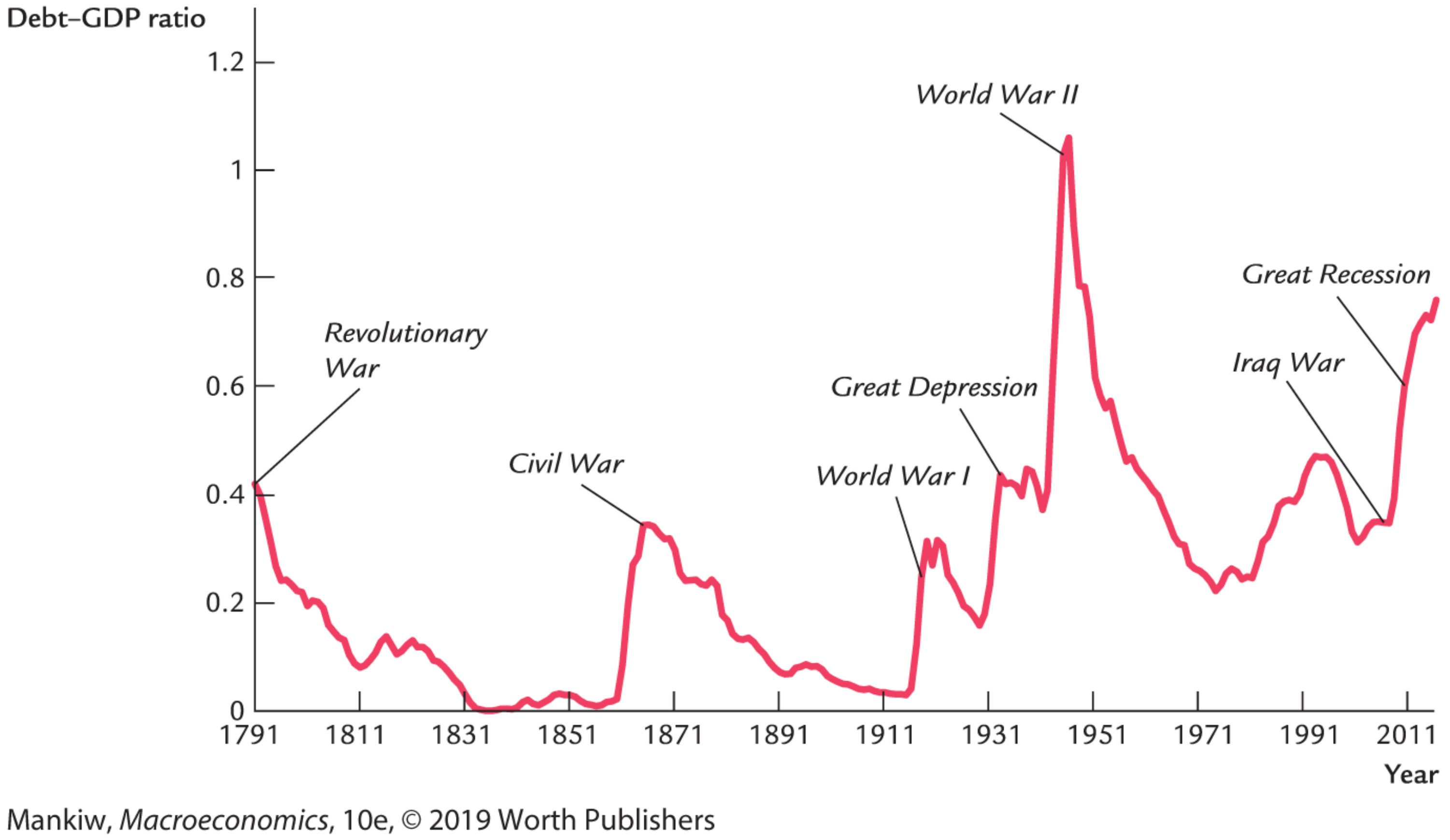


FIGURE 17-1 The Ratio of Government Debt to GDP Since 1791 The U.S. federal government debt held by the public, relative to the size of the U.S. economy, rises sharply during wars, when the government finances wartime spending by borrowing. It also increases during major economic downturns, such as the Great Depression of the 1930s and the Great Recession following the financial crisis of 2008–2009. The debt-to-GDP ratio usually declines gradually during periods of peace and prosperity.

Data from: U.S. Department of the Treasury, U.S. Department of Commerce, and T. S. Berry, “Production and Population Since 1789,” Bostwick Paper No. 6, Richmond, 1988.

Historically, the main cause of increases in the government debt is war. The debt-to-GDP ratio rises sharply during major wars and falls slowly during peacetime. Many economists think that this historical pattern is the appropriate way to run fiscal policy. As we discuss later in the chapter, deficit financing of wars appears optimal to keep tax rates stable and to shift some of the tax burden from current to future generations.

One instance of a large increase in government debt in peacetime began in the early 1980s. When Ronald Reagan was elected president in 1980, he was committed to reducing taxes and increasing military spending. These policies, coupled with a deep recession due to tight monetary policy, began a long period of large budget deficits. The government debt expressed as a percentage of GDP roughly doubled from 25 percent in 1980 to 47 percent in 1995. The United States had never before experienced such a big increase in government debt during a period of peace and prosperity. Many economists criticized these policies as imposing an unjustifiable burden on future generations.

The increase in government debt during the 1980s caused concern among many policymakers as well. In 1990, President George H. W. Bush raised taxes to reduce the deficit, breaking his “Read my lips: No new taxes” campaign pledge and, according to some political commentators, costing him reelection. In 1993, when President Bill Clinton took office, he raised taxes yet again. These tax increases, together with spending restraint and rapid economic growth due to the information-technology boom, caused the budget deficits to shrink and eventually turn into budget surpluses. The government debt fell from 47 percent of GDP in 1995 to 31 percent in 2001.

When President George W. Bush took office in 2001, the high-tech boom in the stock market was reversing course, and the economy was heading into recession. Economic downturns automatically cause tax revenue to fall and push the budget toward deficit. In addition, tax cuts to combat the recession and increased spending for homeland security and wars in Afghanistan and Iraq further increased the budget deficit. From 2001 to 2008, government debt rose from 31 to 39 percent of GDP.

When President Barack Obama moved into the White House in 2009, the economy was in the midst of the Great Recession. Tax revenues declined as the economy shrank. In addition, as discussed in Chapter 11, President Obama signed into law a fiscal stimulus to prop up aggregate demand. As a result, the government ran large budget deficits, and the debt-to-GDP ratio rose to 70 percent of GDP by 2012.

These trends led to a notable event in August 2011: Standard & Poor’s, a private agency that evaluates the safety of bonds, reduced its credit rating on U.S. government debt to one notch below the top AAA grade. For many years, U.S. government debt was considered the safest around. That is, buyers of these bonds could be confident that they would be repaid in full when the bond matured. Standard & Poor’s, however, was sufficiently concerned about fiscal policy that it raised the possibility that the U.S. government might someday default.

As the economy recovered from the Great Recession, the budget deficit shrank, and the rise in the debt-to-GDP ratio slowed. In 2016, when President Donald Trump was elected, the federal debt was 76 percent of GDP. President Trump’s first major economic initiative was a cut in taxes, especially on corporate income, effective in 2018. Supporters of the policy thought it would promote capital accumulation and economic growth, whereas critics believed it would result in an excessive increase in government debt.

As this book was going to press, Standard & Poor’s had not changed its rating on U.S. government debt. Explaining its decision, it cited the nation’s economic strength on the one hand and its “high level of debt and weak political cohesion” on the other.

CASE STUDY

The Troubling Long-Term Outlook for Fiscal Policy

What does the future hold for budget deficits and government debt? When economists project the path of U.S. fiscal policy over the next several decades, they paint a troubling picture.

One reason is demographic. Advances in medical technology have been increasing life expectancy, while improvements in birth control and changing social norms have reduced the number of children people have. Because of these developments, the elderly are becoming a larger share of the population. In 1950, the elderly (aged 65 and older) made up only 8 percent of the overall population. By 2016, the elderly’s share had risen to 15 percent, and it is expected to rise to about 20 percent in 2050. More than a third of the federal budget is now devoted to providing the elderly with pensions (through Social Security) and health care (through Medicare). As more people become eligible for these programs over time, government spending automatically increases.

A second, related reason for the troubling fiscal picture is the rising cost of health care. The government provides health care to the elderly through Medicare and to the poor through Medicaid, and since the passage of the Affordable Care Act in 2010, it also subsidizes private health insurance for low-income households. As the cost of health care increases, government spending on these programs increases as well. Policymakers have proposed various ways to stem the rise of health care costs, such as reducing the burden of lawsuits, encouraging more competition among health care providers, promoting greater use of information technology, and reducing unnecessary testing and treatment by changing how physicians are paid. Yet many health economists believe such measures will have limited impact. The main reason for rising health care costs is medical advances that provide new and better but often expensive ways to extend and improve our lives.

The combination of the aging population and rising health care costs will have a major impact on the federal budget. Government spending on Social Security and on Medicare, Medicaid, and other health care programs has already risen from less than 1 percent of GDP in 1950 to 10.4 percent in 2017. The Congressional Budget Office estimates that if no changes are made, spending on these programs will rise to 15.5 percent of GDP by 2047.

How the United States will handle these spending pressures is an open question. The key issue is how the required fiscal adjustment will be split between tax increases and spending reductions. Some economists believe that to pay for these commitments, we will need to raise taxes as a percentage of GDP substantially above historical levels. Others believe that such high tax rates would impose too great a cost on younger workers. They suggest that policymakers should reduce the promises now being made to the elderly of the future and that people should be encouraged to take a greater role in providing for themselves as they age. For example, this might entail raising the normal retirement age, while encouraging people to save more during their working years so they can be responsible for more of their own retirement and health care costs.

Resolving this debate will be one of the great policy challenges in the decades ahead. Neither large tax hikes nor large spending cuts are politically popular, which is why the problem has not been addressed already. Yet the only alternative is a continuation of large budget deficits and increasing government debt. At some point, as government debt rises as a share of GDP, the government’s ability or willingness to service and repay these debts would be called into question. ■

17-2 Measurement Problems

The government budget deficit equals government spending minus government revenue, which is the amount of new debt the government needs to issue to finance its operations. Although this definition sounds simple, debates about fiscal policy sometimes arise over how the budget deficit should be measured. Some economists believe that the deficit as conventionally measured is not a good indicator of the stance of fiscal policy. That is, they believe that the budget deficit does not accurately gauge either the impact of fiscal policy on today's economy or the burden being placed on future generations of taxpayers. In this section we discuss four problems with the usual measure of the budget deficit.

Problem 1: Inflation

The least controversial of the measurement problems is the correction for inflation. Almost all economists agree that the government's indebtedness should be measured in real terms, not in nominal terms. The measured deficit should equal the change in the government's real debt, not the change in its nominal debt.

The budget deficit as commonly measured, however, does not correct for inflation. To see how large an error this induces, consider the following example. Suppose the real government debt is not changing; in other words, in real terms, the budget is balanced. In this case, the nominal debt must be rising at the rate of inflation. That is,

$$\Delta D / D = \pi,$$

where π is the inflation rate and *D* is the stock of government debt. This implies

$$\Delta D = \pi D.$$

The government would look at the change in the nominal debt ΔD and would report a budget deficit of πD. Hence, most economists believe that the reported budget deficit is overstated by the amount πD.

We can make the same argument in another way. The deficit is government expenditure minus government revenue. Part of expenditure is the interest paid on the government debt. Expenditure should include only the real interest paid on the debt *rD*, not the nominal interest paid *iD*. Because the difference between the nominal interest rate *i* and the real interest rate *r* is the inflation rate π, the budget deficit is overstated by πD.

This correction for inflation can be large when inflation is high, and it can often change our evaluation of fiscal policy. For example, in 1979, the federal government reported a budget deficit of $28 billion. Inflation was 8.6 percent, and the government debt held by the public at the beginning of the year was $495 billion. The deficit was therefore overstated by

$$\begin{aligned} \pi D &= 0.086 \times \$495 \text{ billion} \\ &= \$43 \text{ billion.} \end{aligned}$$

Corrected for inflation, the reported budget deficit of $28 billion turns into a budget surplus of $15 billion! In other words, even though nominal government debt was rising, real government debt was falling. This correction has been less important in recent years because inflation has been low.

Problem 2: Capital Assets

Many economists believe that an accurate assessment of the government's budget deficit requires taking into account the government's assets as well as its liabilities. In particular, when measuring the government's overall indebtedness, we should subtract government assets from government debt. Therefore, the budget deficit should be measured as the change in debt minus the change in assets.

Certainly, individuals and firms treat assets and liabilities symmetrically. When a person borrows to buy a house, we do not say that she is running a budget deficit. Instead, we offset the increase in assets (the house) against the increase in debt (the mortgage) and record no change in net wealth. Perhaps we should treat the government's finances the same way.

A budget procedure that accounts for both assets and liabilities is called **capital budgeting** because it takes into account changes in capital. For example, suppose the government sells one of its office buildings or some of its land and uses the proceeds to reduce the government debt. Under current budget procedures, the reported deficit would be lower. Under capital budgeting, the revenue received from the sale would not lower the deficit because the reduction in debt would be offset by a reduction in assets. Similarly, under capital budgeting, government borrowing to finance the purchase of a capital good would not raise the deficit.

The major difficulty with capital budgeting is that it is hard to decide which government expenditures should count as capital expenditures. For example, should the interstate highway system be counted as an asset of the government? If so, what is its value? What about the stockpile of nuclear weapons? Should spending on education be treated as expenditure on human capital? These difficult questions must be answered if the government adopts a capital budget.

Reasonable people disagree about whether the federal government should use capital budgeting. (Many state governments already use it.) Opponents of capital budgeting argue that, although the system is superior in principle to the current system, it is too difficult to implement. Proponents argue that even an imperfect treatment of capital assets would be better than ignoring them altogether.

Problem 3: Uncounted Liabilities

Some economists argue that the measured budget deficit is misleading because it excludes some important government liabilities. For example, consider the pensions of government workers. These workers provide labor services to the government today, but part of their compensation is deferred to the future. In essence, these workers are providing a loan to the government. Their future pension benefits are a government liability similar to government debt. Yet this liability is not included in the government debt, and the accumulation of this liability is not included in the budget deficit. According to some estimates, this implicit liability is almost as large as the official government debt.

Similarly, consider the Social Security system. In some ways, the system is like a pension plan. People pay some of their income into the system when young and expect to receive benefits when old. Perhaps accumulated future Social Security benefits should also be included in the government's liabilities. Estimates suggest that the government's future Social Security liabilities (less future Social Security taxes) are more than triple the government debt as officially measured.

One might argue that Social Security liabilities are different from government debt because the government can change the laws determining Social Security benefits. Yet, in principle, the government could always choose not to repay all of its debt: the government honors its debt only because it chooses to do so. Promises to pay the holders of government debt may not be fundamentally different from promises to pay the future recipients of Social Security.

A particularly difficult-to-measure form of government liability is the *contingent liability*—the liability that is due only if a specified event occurs. For example, the government guarantees many forms of private credit, such as student loans, mortgages for low- and moderate-income families, and deposits in banks and savings-and-loan institutions. If the borrower repays the loan, the government pays nothing; if the borrower defaults, the government makes the repayment. When the government provides this guarantee, it undertakes a liability contingent on the borrower's default. Yet this contingent liability is not reflected in the budget deficit, in part because it is unclear what dollar value to attach to it.

Problem 4: The Business Cycle

Many changes in the government's budget deficit occur automatically in response to a fluctuating economy. When the economy goes into a recession, incomes fall, so people pay less in personal income taxes. Profits fall, so corporations pay less in corporate income taxes. Fewer people are employed, so payroll tax revenue declines. More people become eligible for government assistance, such as welfare and unemployment insurance, so government spending rises. Even without any change in the laws governing taxation and spending, the budget deficit increases.

These automatic changes in the deficit are not errors in measurement because the government truly borrows more when a recession depresses tax revenue and boosts government spending. But these changes make it harder to use the deficit to monitor changes in fiscal policy. That is, the deficit can rise or fall either because the government has changed policy or because the economy has changed direction. For some purposes, it would be good to know which is occurring.

To solve this problem, the government calculates a **cyclically adjusted budget deficit** (sometimes called the *full-employment budget deficit*). The cyclically adjusted deficit is based on estimates of what government spending and tax revenue would be if the economy were operating at its natural level of output and employment. The cyclically adjusted deficit is a useful measure because it reflects policy changes but not the current stage of the business cycle.

Summing Up

Economists differ in the importance they place on these measurement problems. Some believe that the problems are so severe that the budget deficit as normally measured is almost meaningless. Most take these measurement problems seriously but still view the measured budget deficit as a useful indicator of fiscal policy.

The undisputed lesson is that to fully evaluate fiscal policy, economists and policymakers must look at more than just the measured budget deficit. And, in fact, they do. The budget documents prepared annually by the Office of Management and Budget contain detailed information about the government's finances, including data on capital expenditures and credit programs.

No economic statistic is perfect. Whenever we see a number reported in the media, we need to know what it is measuring and what it is leaving out. This is especially true for data on government debt and budget deficits.

17-3 The Traditional View of Government Debt

Imagine that you are an economist working for the Congressional Budget Office (CBO). You receive a letter from the chair of the Senate Budget Committee:

> Dear CBO Economist:
>
> Congress is considering the president's request to cut all taxes by 20 percent, and my committee would like your analysis. We see little hope of reducing government spending, so the tax cut would increase the budget deficit. How would the tax cut and budget deficit affect the economy and the economic well-being of the country?
>
> Sincerely,
>
> Committee Chair

Before responding to the senator, you open your favorite economics textbook—this one, of course—to see what the models predict.

To analyze the long-run effects of this policy change, you turn to the models in Chapters 3 through 9. The model in Chapter 3 shows that a tax cut stimulates consumer spending and reduces national saving. The reduction in saving raises the interest rate, crowding out investment. The Solow growth model, introduced in Chapter 8, shows that lower investment eventually leads to a lower steady-state capital stock and a lower level of output. Because we concluded in Chapter 9 that the U.S. economy has less capital than in the Golden Rule steady state (the steady state with maximum consumption), the fall in steady-state capital means lower consumption and reduced economic well-being.

To analyze the short-run effects of the policy change, you turn to the *IS*–*LM* model in Chapters 11 and 12. This model shows that a tax cut stimulates consumer spending, which implies an expansionary shift in the *IS* curve. If there is no change in monetary policy, the shift in the *IS* curve leads to an expansionary shift in the aggregate demand curve. In the short run, when prices are sticky, the expansion in aggregate demand leads to higher output and lower unemployment. Over time, as prices adjust, the economy returns to the natural level of output, and the higher aggregate demand results in a higher price level.

To see how international trade affects your analysis, you turn to the open-economy models in Chapters 6 and 13. The model in Chapter 6 shows that when national saving falls, people start financing investment by borrowing from abroad, causing a trade deficit. Although the inflow of capital from abroad mitigates the effect of the fiscal-policy change on U.S. capital accumulation, the United States becomes indebted to foreign countries. The fiscal-policy change also causes the dollar to appreciate, making foreign goods cheaper in the United States and domestic goods more expensive abroad. The Mundell–Fleming model in Chapter 13 shows that the appreciation of the dollar and the resulting fall in net exports reduce the short-run expansionary impact of the fiscal change on output and employment.

With these models in mind, you draft a response:

Dear Senator:

A tax cut financed by government borrowing would have many effects on the economy. The immediate impact of the tax cut would be to stimulate consumer spending. Higher consumer spending affects the economy in both the short run and the long run.

In the short run, higher consumer spending would raise the demand for goods and services and thus raise output and employment. Interest rates would also rise, however, as investors competed for a smaller flow of saving. Higher interest rates would discourage investment and would encourage capital to flow in from abroad. The dollar would appreciate against foreign currencies, making U.S. firms less competitive in world markets.

In the long run, the smaller national saving resulting from the tax cut would mean a smaller capital stock and a greater foreign debt. Therefore, the output of the nation would be smaller, and a greater share of that output would be owed to foreigners.

The overall effect of the tax cut on economic well-being is hard to judge. Current generations would benefit from higher consumption and higher employment, although inflation would likely be higher as well. Future generations would bear much of the burden of today's budget deficits: they would be born into a nation with a smaller capital stock and a larger foreign debt.

Your faithful servant,

CBO Economist

The senator replies:

Dear CBO Economist:

Thank you for your letter. It made sense to me. But yesterday my committee heard testimony from a "Ricardian" economist who reached a different conclusion. She said that a tax cut by itself would not stimulate consumer spending. Thus, she concluded that the budget deficit would not have all the effects you listed. What’s going on here?

Sincerely,

Committee Chair

After studying the next section, you write back to the senator, explaining in detail the debate over Ricardian equivalence.

FYI

Taxes and Incentives

Throughout this book we have summarized the tax system with a single variable, *T*. In our models, the policy instrument is the level of taxation that the government chooses; we have ignored the issue of how the government raises this tax revenue. In practice, however, taxes are not lump-sum payments but are levied on some type of economic activity. The U.S. federal government raises revenue from personal income taxes (47 percent of tax revenue in 2016), payroll taxes (34 percent), corporate profits taxes (9 percent), and several other sources (9 percent).

Courses in public finance spend much time studying the pros and cons of alternative taxes. One lesson emphasized in such courses is that taxes affect incentives. When people are taxed on their labor earnings, they have less incentive to work hard. When people are taxed on the income from owning capital, they have less incentive to save and invest in capital. As a result, when taxes change, incentives change, and this can have macroeconomic effects. If lower tax rates encourage increased work and investment, the aggregate supply of goods and services increases.

Some economists, called *supply-siders*, believe that the incentive effects of taxes are large. Some supply-siders even suggest that tax cuts can be self-financing: a cut in tax rates induces such a large increase in aggregate supply that tax revenue increases, despite the fall in tax rates. Although all economists agree that taxes affect incentives and that incentives affect aggregate supply to some degree, most believe that the incentive effects are not large enough to make tax cuts self-financing in most circumstances.[1]

17-4 The Ricardian View of Government Debt

The traditional view of government debt presumes that when the government cuts taxes and runs a budget deficit, consumers respond to their higher after-tax income by spending more. An alternative view, called **Ricardian equivalence**, questions this presumption. According to the Ricardian view, consumers are forward-looking and, therefore, base their spending decisions not only on their current income but also on their expected future income. As we will see in Chapter 19, the forward-looking consumer is at the heart of many modern theories of consumption. The Ricardian view of government debt applies the logic of the forward-looking consumer to analyzing fiscal policy.

The Basic Logic of Ricardian Equivalence

Consider how a forward-looking consumer responds to the tax cut that the Senate Budget Committee is considering. The consumer might reason as follows:

> The government is cutting taxes without any plans to reduce government spending. Does this policy alter my set of opportunities? Am I richer because of this tax cut? Should I consume more?
>
> Maybe not. The government is financing the tax cut by running a budget deficit. At some point in the future, the government will have to raise taxes to pay off the debt and accumulated interest. So the policy really involves a tax cut today coupled with a tax hike in the future. The tax cut merely gives me transitory income that eventually will be taken back. I am not any better off, so I will leave my consumption unchanged.

The forward-looking consumer understands that government borrowing today means higher taxes in the future. A tax cut financed by government debt does not reduce the tax burden; it merely reschedules it. It therefore should not encourage the consumer to spend more.

One can view this argument another way. Suppose the government borrows $1,000 from the typical citizen to give that citizen a $1,000 tax cut. In essence, this policy is the same as giving the citizen a $1,000 government bond as a gift. One side of the bond says, “The government owes you, the bondholder, $1,000 plus interest.” The other side says, “You, the taxpayer, owe the government $1,000 plus interest.” Overall, the gift of a bond from the government to the typical citizen does not make the citizen richer or poorer because the value of the bond is offset by the value of the future tax liability.

The general principle is that government debt is equivalent to future taxes, and if consumers are sufficiently forward-looking, future taxes are equivalent to current taxes. Hence, financing the government by debt is equivalent to financing it by taxes. This view is called *Ricardian equivalence*, after the nineteenth-century economist David Ricardo, who first noted this theoretical argument.

The implication of Ricardian equivalence is that a debt-financed tax cut does not affect consumption. Households save the extra disposable income to pay the future tax liability that the tax cut implies. This increase in private saving offsets the decrease in public saving. National saving—the sum of private and public saving—remains the same. Hence, the tax cut has none of the effects that the traditional analysis predicts.

The logic of Ricardian equivalence does not mean that all changes in fiscal policy are irrelevant. Changes in fiscal policy influence consumer spending if they influence present or future government purchases. For example, suppose the government cuts taxes today because it plans to reduce government purchases in the future. If the consumer understands that this tax cut does not require an increase in future taxes, she feels richer and raises her consumption. But it is the reduction in government purchases, not the reduction in taxes, that stimulates consumption: the announcement of a future reduction in government purchases would raise consumption today even if current taxes were unchanged because it would imply lower taxes at some time in the future.

Consumers and Future Taxes

The essence of the Ricardian view is that when people decide how much to consume, they rationally look ahead to the future taxes implied by government debt. But how forward-looking are consumers? Defenders of the traditional view of government debt believe that the prospect of future taxes does not have as large an influence on current consumption as the Ricardian view assumes. Here are some of their arguments.[2]

Myopia

Proponents of the Ricardian view of fiscal policy assume that people are rational when deciding how much of their income to consume and how much to save. When the government borrows to pay for current spending, rational consumers look ahead to the future taxes required to support this debt. Thus, the Ricardian view presumes that people have substantial knowledge and foresight.

One argument for the traditional view of tax cuts is that people are shortsighted, perhaps because they do not fully comprehend the implications of government budget deficits. It is possible that some people follow simple and not fully rational rules of thumb when choosing how much to save. Suppose, for example, a person assumes that future taxes will be the same as current taxes. This person will ignore future changes in taxes required by current government policies. A debt-financed tax cut will lead this person to believe that her lifetime income has increased, even if it hasn't. The tax cut will increase consumption and lower national saving.

Borrowing Constraints

The Ricardian view of government debt assumes that consumers base their spending not on their current income but on their lifetime income, which includes both current and expected future income. According to the Ricardian view, a debt-financed tax cut increases current income, but it does not alter lifetime income or consumption. Advocates of the traditional view of government debt argue that current income is more important than lifetime income for consumers who face binding borrowing constraints. A *borrowing constraint* is a limit on how much an individual can borrow from banks or other financial institutions.

A person who would like to consume more than her current income allows—perhaps because she expects higher income in the future—has to do so by borrowing. If she cannot borrow to finance current consumption, or if she can borrow only a limited amount, her current income determines her spending, regardless of what her lifetime income might be. In this case, a debt-financed tax cut raises current income and thus consumption, even though future income will be lower. In essence, when the government cuts current taxes and raises future taxes, it is giving taxpayers a loan. If a person wanted to obtain a loan but was unable to, the tax cut expands her opportunities and stimulates consumption.

CASE STUDY

George H. W. Bush's Withholding Experiment

In early 1992, President George H. W. Bush pursued a novel policy to deal with the lingering recession in the United States. By executive order, he lowered the amount of income taxes that were being withheld from workers' paychecks. The order did not reduce the amount of taxes that workers owed; it merely delayed payment. The higher take-home pay that workers received during 1992 was to be offset by higher tax payments, or smaller tax refunds, when income taxes were due in April 1993.

What effect would you predict for this policy? According to the logic of Ricardian equivalence, consumers should have realized that their lifetime resources were unchanged and, therefore, saved the extra take-home pay to meet the upcoming tax liability. Yet George Bush claimed his policy would provide "money people can use to help pay for clothing, college, or to get a new car." That is, he believed that consumers would spend the extra income, thereby stimulating aggregate demand and helping the economy recover from the recession. Bush seemed to assume that consumers were shortsighted or faced binding borrowing constraints.

Gauging the actual effects of this policy is difficult with aggregate data because many other things were happening at the same time. Yet some evidence comes from a survey two economists conducted shortly after the policy was announced. The survey asked people what they would do with the extra income. Fifty-seven percent of the respondents said they would save it, use it to repay debts, or adjust their withholding in order to reverse the effect of Bush's executive order. Forty-three percent said they would spend the extra income. Thus, for this policy change, a majority of the population was planning to act as Ricardian theory posits. Nonetheless, Bush was partly right: many people planned to spend the extra income, even though they understood that the following year's tax bill would be higher.[3] ■

Future Generations

Besides myopia and borrowing constraints, a third argument for the traditional view of government debt is that consumers expect the implied future taxes to fall not on them but on future generations. Suppose, for example, the government cuts taxes today, issues 30-year bonds to finance the budget deficit, and then raises taxes in 30 years to repay the loan. In this case, the government debt represents a transfer of wealth from the next generation of taxpayers (who face the tax hike) to the current generation of taxpayers (who receive the tax cut). This transfer raises the lifetime resources of the current generation, increasing their consumption. In essence, a debt-financed tax cut stimulates consumption because it gives the current generation the opportunity to consume at the expense of the next generation.

The Wall Street Journal/Cartoon Features Syndicate

"What's this I hear about you adults mortgaging my future?"

Economist Robert Barro has provided a clever rejoinder to this argument to support the Ricardian view. Barro argues that because future generations are the children and grandchildren of the current generation, we should not view these various generations as independent economic actors. Instead, he claims, the appropriate assumption is that current generations care about future generations. This altruism between generations is evidenced by the gifts that people give their children, often in the form of bequests at the time of their deaths. The existence of bequests suggests that many people are not eager to take advantage of the opportunity to consume at their children's expense.

According to Barro's analysis, the relevant decisionmaking unit is not the individual, whose life is finite, but the family, which continues forever. In other words, an individual decides how much to consume based not only on her own income but also on the income of future members of her family. A debt-financed tax cut may raise the income an individual receives in her lifetime, but it does not raise her family's overall resources. Instead of consuming the extra income from the tax cut, the individual saves it and leaves it as a bequest to her children, who will bear the future tax liability.

We can now see that the debate over government debt is really a debate over consumer behavior. The Ricardian view assumes that consumers have a long time horizon. Barro's analysis of the family implies that the consumer's time horizon, like the government's, is effectively infinite. Yet it is possible that consumers do not look ahead to the tax liabilities of future generations. Perhaps they expect their children to be richer than they are and welcome the opportunity to consume at their children's expense. The fact that many people leave zero or minimal bequests to their children is consistent with this hypothesis. For these zero-bequest families, a debt-financed tax cut alters consumption by redistributing wealth among generations.[4]

Making a Choice

Having seen the traditional and Ricardian views of government debt, you should consider two sets of questions.

First, with which view do you agree? If the government cuts taxes today, runs a budget deficit, and raises taxes in the future, how will the policy affect the economy? Will it stimulate consumption, as the traditional view holds? Or will consumers understand that their lifetime income is unchanged and, therefore, offset the budget deficit with higher private saving?

Second, why do you hold the view that you do? If you agree with the traditional view of government debt, what is the reason? Do consumers fail to understand that higher government borrowing today means higher taxes tomorrow? Or do they ignore future taxes either because they face borrowing constraints or because future taxes will fall on future generations with which they do not feel an economic link? If you hold the Ricardian view, do you believe that consumers have the foresight to see that government borrowing today will result in future taxes levied on them or their descendants? Do you believe that consumers will save the extra income to offset that future tax liability?

We might hope that the evidence could help us decide between these two views of government debt. Yet when economists examine historical episodes of large budget deficits, the evidence is inconclusive.

FYI

Ricardo on Ricardian Equivalence

David Ricardo was a millionaire stockbroker and one of the greatest economists of all time. His most important contribution to the field was his 1817 book *On the Principles of Political Economy and Taxation*, in which he developed the theory of comparative advantage, which economists still use to explain the gains from international trade. Ricardo was also a member of the British Parliament, where he put his own theories to work and opposed the corn laws, which restricted international trade in grain.

Ricardo was interested in the alternative ways a government might pay for its expenditure. In an 1820 article called "Essay on the Funding System," he considered an example of a war that cost 20 million pounds. He noted that if the interest rate was 5 percent, this expense could be financed with a one-time tax of 20 million pounds, a perpetual tax of 1 million pounds, or a tax of 1.2 million pounds for 45 years. He wrote:

> In point of economy there is no real difference in either of the modes, for 20 million in one payment, 1 million per annum forever, or 1,200,000 pounds for forty-five years, are precisely of the same value.

Ricardo was aware that the issue involved the linkages among generations:

> It would be difficult to convince a man possessed of 20,000 pounds, or any other sum, that a perpetual payment of 50 pounds per annum was equally burdensome with a single tax of 1000 pounds. He would have some vague notion that the 50 pounds per annum would be paid by posterity, and would not be paid by him; but if he leaves his fortune to his son, and leaves it charged with this perpetual tax, where is the difference whether he leaves him 20,000 pounds with the tax, or 19,000 pounds without it?

Although Ricardo viewed these alternative methods of government finance as equivalent, he did not think other people would view them as such:

> The people who pay the taxes . . . do not manage their private affairs accordingly. We are too apt to think that the war is burdensome only in proportion to what we are at the moment called to pay for it in taxes, without reflecting on the probable duration of such taxes.

Thus, Ricardo doubted that people were rational and farsighted enough to look ahead fully to their future tax liabilities.

As a policymaker, Ricardo took the government debt seriously. Before the British Parliament, he once declared:

> This would be the happiest country in the world, and its progress in prosperity would be beyond the power of imagination to conceive, if we got rid of two great evils—the national debt and the corn laws.

It is one of the great ironies in the history of economic thought that Ricardo rejected the theory that now bears his name!

Consider, for example, the experience of the 1980s. The large budget deficits, caused partly by the Reagan tax cut of 1981, seem to offer a natural experiment to test the two views of government debt. At first glance, this episode appears to support the traditional view. The large budget deficits coincided with low national saving, high real interest rates, and a large trade deficit. Advocates of the traditional view of government debt often claim that this experience confirms their position.

Yet those who hold the Ricardian view of government debt interpret these events differently. Perhaps saving was low in the 1980s because people were optimistic about future growth—an optimism that was also reflected in a booming stock market. Or perhaps saving was low because people expected that the tax cut would eventually lead not to higher taxes but, as Reagan promised, to lower government spending. Because it is hard to rule out any of these interpretations, both views of government debt survive.

17-5 Other Perspectives on Government Debt

Policy debates over government debt have many facets. So far we have considered the traditional and Ricardian views of government debt. According to the traditional view, a government budget deficit expands aggregate demand and stimulates output in the short run but crowds out capital and depresses growth in the long run. According to the Ricardian view, a government budget deficit has none of these effects because consumers understand that a budget deficit merely represents the postponement of a tax burden. With these two theories as background, we now consider several other perspectives on government debt.

Balanced Budgets Versus Optimal Fiscal Policy

In the United States, many state constitutions require the state government to run a balanced budget. A recurring political debate is whether the U.S. Constitution should require a balanced budget for the federal government as well. Most economists oppose a strict rule requiring the government to balance its budget. There are three reasons that optimal fiscal policy may at times call for a budget deficit or surplus.

Stabilization

A budget deficit or surplus can help stabilize the economy. In essence, a balanced-budget rule would revoke the automatic stabilizing powers of the system of taxes and transfers. When the economy goes into a recession, taxes automatically fall, and transfers automatically rise. These automatic responses help stabilize the economy, but they push the budget into deficit. A strict balanced-budget rule would require that the government raise taxes or reduce spending in a recession, which would depress aggregate demand and deepen the downturn.

Tax Smoothing

A budget deficit or surplus can be used to reduce the distortion of incentives caused by the tax system. As discussed earlier, high tax rates impose a cost on society by discouraging economic activity. A tax on labor earnings, for instance, reduces the incentive that people have to work long hours. Because this disincentive becomes particularly large at very high tax rates, the total social cost of taxes is minimized by keeping tax rates stable rather than making them high in some years and low in others. Economists call this policy **tax smoothing**. To keep tax rates smooth, a deficit is necessary in years of unusually low income (recessions) or unusually high expenditure (wars).

Intergenerational Redistribution

A budget deficit can be used to shift a tax burden from current to future generations. For example, some economists argue that if the current generation fights a war to preserve freedom, future generations also benefit and should bear some of the burden. To pass on some of the war's costs, the current generation can finance the war with a budget deficit. The government can later retire the debt by levying taxes on the next generation.

These considerations lead most economists to reject a strict balanced-budget rule. At the very least, a rule for fiscal policy needs to take account of the recurring episodes, such as recessions and wars, during which it is reasonable for the government to run a budget deficit.

Fiscal Effects on Monetary Policy

In 1985, Paul Volcker told Congress that "the actual and prospective size of the budget deficit . . . heightens skepticism about our ability to control the money supply and contain inflation." A decade later, Alan Greenspan claimed that "a substantial reduction in the long-term prospective deficit of the United States will significantly lower very long-term inflation expectations." Both of these Fed chairs apparently saw a link between fiscal policy and monetary policy.

We first discussed such a possibility in Chapter 5. As we saw, one way for a government to finance a budget deficit is simply to print money—a policy that leads to higher inflation. Indeed, when countries experience hyperinflation, the typical reason is that fiscal policymakers are relying on the inflation tax to pay for some of their spending. The ends of hyperinflations almost always coincide with fiscal reforms that include large cuts in government spending and a reduced need for seigniorage.

In addition to this link between the budget deficit and inflation, some economists have suggested that a high level of debt might encourage the government to create inflation. Because most government debt is specified in nominal terms, the real value of the debt falls when the price level rises. This is the usual redistribution between creditors and debtors caused by unexpected inflation; here the debtor is the government, and the creditor is the private sector. But this debtor, unlike others, has access to the monetary printing press. A high level of debt might encourage the government to print money, thereby raising the price level and reducing the real value of its debts.

Despite these concerns about a possible link between government debt and monetary policy, there is little evidence that this link is important in most developed countries. In the United States, for instance, inflation was high in the 1970s, even though government debt was low relative to GDP. Monetary policymakers got inflation under control in the early 1980s, just as fiscal policymakers started running large budget deficits and increasing the government debt. In 2017, the debt-to-GDP ratio was high by historical standards, but inflation was a bit below the Fed’s announced target of 2 percent. Thus, although monetary policy may sometimes be driven by fiscal policy, such as during classic hyperinflations, this situation is not the norm in most countries today. There are several reasons for this. First, most governments can finance deficits by selling debt and don’t need to rely on seigniorage. Second, central banks often have enough independence to resist political pressure. Third, most policymakers know that inflation is a poor solution to fiscal problems.[5]

Debt and the Political Process

Fiscal policy is made not by angels but by an imperfect political process. Some economists worry that the possibility of financing government spending by issuing debt makes that political process all the worse.

This idea has a long history. Nineteenth-century Swedish economist Knut Wicksell claimed that if the benefit of some type of government spending exceeded its cost, it should be possible to finance that spending in a way that would receive unanimous support from the voters. He concluded that government spending should be undertaken only when support is, in fact, nearly unanimous. In the case of debt finance, however, Wicksell was concerned that “the interests [of future taxpayers] are not represented at all or are represented inadequately in the tax-approving assembly.”

Many economists have echoed this theme more recently. In their 1977 book *Democracy in Deficit*, James Buchanan and Richard Wagner argued for a balanced-budget rule for fiscal policy on the grounds that it “will have the effect of bringing the real costs of public outlays to the awareness of decision makers; it will tend to dispel the illusory ‘something for nothing’ aspects of fiscal choice.” Similarly, Martin Feldstein (once an economic adviser to Ronald Reagan and a long-time critic of budget deficits) argued that “only the ‘hard budget constraint’ of having to balance the budget” can force politicians to judge whether spending’s “benefits really justify its cost.”

These arguments have led some economists to favor a constitutional amendment requiring Congress to pass a balanced budget. Often these proposals have escape clauses for times of national emergency, such as wars and depressions, when a budget deficit is a reasonable policy response. Some critics of these proposals argue that, even with the escape clauses, such a constitutional amendment would tie the hands of policymakers too severely. Others claim that Congress would evade the balanced-budget requirement with accounting tricks. As this discussion makes clear, the debate over the desirability of a balanced-budget amendment is as much political as economic.

International Dimensions

Government debt may affect a nation's role in the world economy. As we first saw in Chapter 6, when a government budget deficit reduces national saving, it often leads to a trade deficit, which in turn is financed by borrowing from abroad. For instance, many observers have blamed U.S. fiscal policy for the switch of the United States from a major creditor to a major debtor in the world economy. This link between the budget deficit and the trade deficit leads to two further effects of government debt.

First, high levels of government debt may increase the risk that an economy will experience capital flight—an abrupt decline in the demand for a country's assets in world financial markets. International investors are aware that a government can always deal with its debt simply by defaulting. This approach was used as far back as 1335, when England's King Edward III defaulted on his debt to Italian bankers. More recently, Russia defaulted on its debts in 1998, and Argentina did the same in 2001. The higher the level of the government debt, the greater the temptation to default. Thus, as government debt increases, international investors may fear default and curtail their lending. If this loss of confidence occurs suddenly, the result could be the classic symptoms of capital flight: a collapse in the value of the currency and an increase in interest rates. As we discussed in Chapter 13, this is what happened to Mexico in the early 1990s when default appeared likely.

Second, high levels of government debt financed by foreign borrowing may reduce a nation's political clout in world affairs. This fear was emphasized by economist Benjamin Friedman in his 1988 book *Day of Reckoning*. He wrote, "World power and influence have historically accrued to creditor countries. It is not coincidental that America emerged as a world power simultaneously with our transition from a debtor nation . . . to a creditor supplying investment capital to the rest of the world." Friedman suggests that if the United States continues to run large trade deficits, it will eventually lose some of its international influence. So far, the record has not been kind to this hypothesis: the United States has run trade deficits throughout the 1980s, 1990s, and the first decade of the 2000s and, nonetheless, remains a leading superpower. But perhaps other events—like the collapse of the Soviet Union—offset the decrease in political clout that the United States would have experienced because of its increased indebtedness.

17-6 Conclusion

Fiscal policy and government debt are central in the political and economic debate worldwide. This chapter discussed some of the issues that lie behind the policy decisions. As we have seen, economists don't always agree about the effects of government indebtedness or about what fiscal policy is best. And, of course, economists are not in charge of designing and enacting fiscal policies. That role goes to our elected leaders, who follow economists' advice only when they choose to.

CHAPTER 18

The Financial System: Opportunities and Dangers

> When written in Chinese, the word “crisis” is composed of two characters—one represents danger and one represents opportunity.
>
> —John F. Kennedy

In 2008 and 2009, the U.S. economy experienced a historic crisis. As we discussed in previous chapters, a decline in house prices led to problems in many financial institutions, causing the most severe economic downturn since the Great Depression of the 1930s. This event was a vivid reminder of the inextricable links between the financial system and the broader economy. When Wall Street sneezes, Main Street catches a cold.

In this chapter we examine the links between the economy and the financial system more thoroughly. We discuss what the financial system is and how it works. We also discuss the new challenges that the financial system presents to policymakers charged with promoting short-run economic stability and long-run economic growth.

The financial system has been present in much of the macroeconomic theory developed throughout this book. In Chapter 3 we discussed a model of the loanable-funds market. There we saw that the interest rate adjusts to balance the supply of loanable funds (derived from the nation’s saving) and the demand for loanable funds (for purposes of investment). In Chapters 8 and 9 we used the Solow model to examine the sources of long-run economic growth. In that model, the financial system is in the background, ensuring that the economy’s saving is directed into investment and capital accumulation.

The financial system has also been present in our short-run analysis. In the *IS–LM* model of Chapters 11 and 12, the interest rate links the goods market with the money market. In that model, the interest rate determines the costs of both holding money and borrowing to fund investment spending. It is therefore the crucial variable through which monetary policy influences the aggregate demand for goods and services.

By studying the financial system in more detail, we can make our analysis of economic growth and fluctuations more nuanced. The financial system is more than a single market for loanable funds, and there are more prices in this system than a single interest rate. Indeed, the complexity of the financial system is sufficiently great that there is an entire subfield of economics, called *finance*, devoted to its study. This chapter focuses on two topics within finance that are crucial to macroeconomics. We start by examining the role of the financial system in the economy. We then consider the causes of financial crises and the policy responses to them.

18-1 What Does the Financial System Do?

Sam is a rational, forward-looking consumer. He earns a good income of $200,000 a year but does not plan to spend all of it this year. He wants to put some of his income aside, perhaps for retirement, a future vacation, college tuition for his newborn son, or just as a precaution to prepare for future uncertainties. The part of his income that he does not spend contributes to the nation's saving.

Ivy is an entrepreneur starting a new business. She has an idea for a doll that she believes would enchant children around the world and therefore be quite profitable. To put her idea into action, she needs some resources: plastics, molds, fabric, sewing machines, and a building to house her manufacturing operation. Ivy's purchases of these capital goods contribute to the nation's investment.

In short, Sam has some income he wants to save, and Ivy has ideas for investments but may not have the funds to pay for them. The solution is obvious: Sam can finance Ivy's venture. The **financial system** is the term for the institutions in the economy that facilitate the flow of funds between savers and investors. That is, the financial system brings people like Sam and people like Ivy together.

Financing Investment

Throughout much of this book, the financial system has been represented as a single market—the market for loanable funds. Those like Sam, who have some income they don't want to consume immediately, bring their saving to this market so they can lend these funds to others. Those like Ivy, who have investment projects they want to undertake, finance these investments by borrowing in this market. In this simple model, a single interest rate adjusts to bring saving and investment into balance.

The actual financial system is more complex than this description. As in the simple model, the main function of the system is to channel resources from savers into various forms of investment. But the system includes numerous mechanisms to facilitate this transfer of resources.

One piece of the financial system is the set of **financial markets** through which households can directly provide resources for investment. Two important financial markets are the market for **bonds** and the market for **stocks**. A bond represents a loan from the bondholder to the firm; a share of stock represents an ownership claim by the shareholder in the firm. That is, a person who buys a bond from, say, Apple Inc. becomes a creditor of the company, while a person who buys newly issued stock from Apple becomes a part owner of the company. (A purchase of stock on a stock exchange, however, represents a transfer of ownership shares from one person to another and does not provide new funds for investment projects.) Raising investment funds by issuing bonds is called **debt finance**, and raising funds by issuing stock is called **equity finance**. Debt and equity are forms of *direct* finance because the saver knows whose investment project his funds are financing.

Another piece of the financial system is the set of **financial intermediaries** through which households can indirectly provide resources for investment. As the term suggests, a financial intermediary stands between the two sides of the market and helps move financial resources toward their best use. Commercial banks are the best-known type of financial intermediary.[1] They take deposits from savers and use these deposits to make loans to those who have investment projects they need to finance. Other examples of financial intermediaries include mutual funds, pension funds, and insurance companies. When an intermediary is involved, the financing is considered *indirect* because the saver is usually unaware of whose investments his funds are financing.

To continue with our example, Sam and Ivy can take advantage of any of these opportunities. If they know each other, Ivy could borrow money directly from Sam and pay him interest on the loan. In this case, she would in effect be selling him a bond. Or Ivy could, in exchange for Sam's money, give him an ownership stake in her new business, and he would enjoy a share of the future profits. In this case, she would be selling him some stock. Or Sam could deposit his savings in a local bank, which in turn could lend the funds to Ivy. In this last case, he would be financing her new venture indirectly: they might never meet, or even know of each other's existence. In all of these cases, Sam and Ivy engage in a mutually advantageous exchange. Sam finds a way to earn a return on his savings, and Ivy finds a way to finance her investment project.

Sharing Risk

Investment is inherently risky. Ivy's new doll might be the next toy craze, or it might be a flop. Like all entrepreneurs, Ivy is starting her venture because she expects it to be profitable, but she cannot be certain of that outcome.

One function of the financial system is to allocate risk. When Ivy sells stock to Sam, she is sharing the risk of her venture with him. If her doll business is profitable, he will enjoy some of the gains. If it loses money, he will share in the losses. Ivy might be eager to share the risk, rather than bear it all herself, because she is **risk averse**. That is, other things equal, she dislikes uncertainty about her future economic outcomes. Sam might be willing to accept some of the risk if the return he expects from this venture is higher than what he would obtain by putting his savings into safer assets. Thus, equity finance provides a way for entrepreneurs and savers to share the risks and returns associated with the entrepreneur's investment ideas.

In addition, the financial system allows savers to reduce their risk by spreading their wealth across many businesses. Sam knows that Ivy's doll venture is risky, so he would be smart to use only some of his savings to buy stock in her business. He could also buy stock from his friend Steve, who is opening an ice-cream store. And he could buy stock in established companies, such as Exxon, Apple, and Facebook. Because the success of Ivy's doll venture is not perfectly correlated with the success of Steve's ice-cream store, or with the profitability of Exxon, Apple, and Facebook, Sam reduces the risk he faces when he spreads his wealth around. Reducing risk by holding many imperfectly correlated assets is called **diversification**.

Various financial institutions facilitate diversification. Among the most important are mutual funds. **Mutual funds** are financial intermediaries that sell shares to savers and use their funds to buy diversified pools of assets. Even a small saver can put, say, $1,000 into a mutual fund and become a part owner of thousands of businesses. Because the fortunes of these many businesses are not perfectly correlated with one another, putting the $1,000 into a mutual fund is less risky than using all of it to buy stock in a single company.

There are limits, however, to how much diversification reduces risk. Some events affect many businesses at the same time. Such risk is called *systematic risk*. In particular, recessions tend to reduce the demand for most products and the profitability of most businesses. Diversification cannot reduce this kind of risk. Yet it can largely eliminate the risks associated with individual businesses, called *idiosyncratic risk*, such as whether Ivy's doll or Steve's ice cream proves popular. For this reason, it is wise for savers like Sam to limit how much they allocate to the stock of any one company.

Dealing with Asymmetric Information

As Sam considers financing Ivy's business venture, one question is paramount: Will her company succeed? If Sam offers her equity financing, he gets a share of future profits, so the fortune of the business is crucial. Debt financing is safer for Sam because debt holders are paid before equity holders, but Ivy's success is still relevant. If the doll business is a failure, Ivy may not be able to repay the loan. That is, she might default. Not only might Sam not get the interest he was promised, but he might also lose his principal (the amount of the loan).

Making matters worse is the fact that Ivy knows more than Sam about herself and her business. Economists use the phrase **asymmetric information** to describe a situation in which one party in a transaction has more relevant information than the other. There are two types of asymmetric information, both of which may affect Sam's decision about financing Ivy's venture.

The first type of asymmetric information concerns *hidden knowledge about attributes*. Is Ivy's doll design one that will have wide appeal, or is it likely to be a niche product? Is the doll market eager for new products, or is it oversaturated? Is Ivy a talented businessperson? Ivy is more likely than Sam to have reliable answers to these questions. This is often the case: entrepreneurs have more information about whether their investment projects are good ones than do those who provide the financing.

In this situation, Sam should worry about the problem of **adverse selection**. As we noted in Chapter 7 in a different context, the term *adverse selection* describes the tendency of people with more information (here, the entrepreneurs) to sort themselves in a way that disadvantages people with less information (here, those providing the financing). In our example, Sam may be concerned that he will be offered opportunities to finance only less desirable business ventures. If Ivy were more confident in her idea, she might try harder to finance it herself, using more of her own savings. The fact that she is asking Sam to provide financing and share some of the risk suggests that perhaps she knows something adverse that he does not know. As a result, Sam has reason to be wary.

The second type of asymmetric information concerns *hidden knowledge about actions.* Once Ivy obtains financing from Sam, she will have many decisions to make. Will she work long hours at the job or cut out early to play tennis with friends? Will she spend the money she has raised in the most profitable way or use it to provide herself with a cushy office and fancy company car? Ivy can promise to act in the best interests of the business, but it will be hard for Sam to verify that she does so because he won't be at the doll factory every day to observe everything she does.

In this case, the problem that arises is **moral hazard**, the risk that an imperfectly monitored agent will act in a dishonest or otherwise inappropriate way. In particular, entrepreneurs investing other people's money may not look after the investment projects as carefully as those who invest their own money. Once Ivy has Sam's money in hand, she may be tempted to choose the easy life. If she succumbs to moral hazard, she will reduce the firm's profitability and increase the risk that it defaults on its debts.

The financial system includes various institutions to mitigate the effects of adverse selection and moral hazard. Banks are among the most important. When a person applies for a bank loan, he fills out an application that asks detailed questions about his business plan, employment background, credit history, criminal record, and other financial and personal characteristics. Because the application is scrutinized by loan officers trained to evaluate businesses, the bank stands a good chance of uncovering the hidden attributes that lead to adverse selection. In addition, to reduce the problem of moral hazard, bank loans may entail restrictions on how the loan proceeds are spent, and the loan officers may monitor the business after the loan is made. As a result, rather than making a direct loan to Ivy, it may make sense for Sam to deposit his money in a bank, which will lend it to entrepreneurs like Ivy. The bank charges a fee for serving as an intermediary, reflected in the difference between the interest rate it charges on loans and the interest rate it pays on deposits. The bank earns its fee by reducing the problems associated with asymmetric information.

Fostering Economic Growth

In Chapters 8 and 9 we used the Solow model to examine the forces that govern long-run economic growth. In that model, we saw that a nation's saving determines the steady-state levels of capital and income per person. The more a nation saves, the more capital its labor force has to work with, the more it produces, and the more income its citizens enjoy.

The Solow model makes the simplifying assumption that there is only a single type of capital, but the real world includes thousands of firms with diverse investment projects competing for the economy's limited resources. Sam's saving can finance Ivy's doll business, but it could instead finance Steve's ice-cream store, a Boeing aircraft factory, or a Walmart retail outlet. The financial system has the job of allocating the economy's scarce saving among the alternative investments.

Ideally, to allocate saving to investment, all the financial system needs are market forces and the magic of Adam Smith's invisible hand. Firms with particularly productive and profitable investment opportunities will be willing to pay higher interest rates for loans than those with less desirable projects. Thus, if the interest rate adjusts to balance the supply and demand for loanable funds, the economy's saving will be allocated to the best of the many possible investments.

Yet, as we have seen, because the financial system is impeded by asymmetric information, it can deviate from this simple ideal. Banks mitigate adverse selection and moral hazard, but they do not eliminate these problems. As a result, some valuable investment projects may not be undertaken because entrepreneurs cannot raise the funds to finance them. If the financial system fails to allocate the economy's saving to its best uses, the economy's productivity will be lower than it could be.

Government policy can help ensure that the financial system works well. First, it can reduce the problem of moral hazard by prosecuting fraud and similar malfeasance. The law cannot ensure that Ivy will put Sam's money to its best use, but if she uses it to pay her personal living expenses, she may go to jail. Second, the government can reduce the problem of adverse selection by requiring some kinds of disclosure. If Ivy's doll business grows large enough to issue stock on a public stock exchange, the government's Securities and Exchange Commission will require that she release regular reports on her firm's earnings and assets and that these reports be certified by accredited accountants.

Because the quality of legal institutions varies around the world, some countries have better financial systems than others, and this difference is one source of international variation in living standards. Rich nations tend to have larger stock markets and larger banking systems (relative to the size of their economies) than poorer nations. Sorting out cause and effect is difficult when comparing countries. Nonetheless, many economists believe that one reason poor nations remain poor is that their financial systems are unable to direct saving to the best possible investments. These nations can foster economic growth by reforming their legal institutions with an eye toward improving their financial systems. If they succeed, entrepreneurs with good ideas will find it easier to start businesses.

18-2 Financial Crises

So far in this chapter we have discussed how the financial system works. Let's now examine why the financial system might stop working and the macroeconomic ramifications of such a disruption.

When we studied business-cycle theory in Chapters 10 to 14, we saw that many kinds of shocks can lead to short-run fluctuations. A shift in consumer or business confidence, a rise or fall in world oil prices, or a sudden change in monetary or fiscal policy can alter aggregate demand or aggregate supply (or both). When this occurs, output and employment are pushed away from their natural levels, and inflation rises or falls.

Here we focus on one particular kind of shock. A **financial crisis** is a major disruption in the financial system that impedes the economy's ability to intermediate between those who want to save and those who want to borrow and invest. Not surprisingly, given the financial system's central role, financial crises have a broad macroeconomic impact. Throughout history, many of the deepest recessions have followed problems in the financial system. These downturns include the Great Depression of the 1930s and the Great Recession of 2008–2009.

The Anatomy of a Crisis

Financial crises are not all alike, but they share some common features. In a nutshell, here are the six elements that are at the center of most financial crises. The financial crisis of 2008–2009 provides a good example of each element.

1. Asset-Price Booms and Busts

Often a period of optimism, leading to a large increase in asset prices, precedes a financial crisis. Sometimes people bid up the price of an asset above its fundamental value (the true value based on an objective analysis of the cash flows the asset will generate). In this case, the market for that asset is said to be in the grip of a **speculative bubble**. Later, when sentiment shifts and optimism turns to pessimism, the bubble bursts, and prices begin to fall. The decline in asset prices is the catalyst for the financial crisis.

FYI

The Efficient Markets Hypothesis Versus Keynes's Beauty Contest

After a company issues equity, its shares are bought and sold on stock exchanges, where prices are set by supply and demand. One continuing debate among economists is whether fluctuations in stock prices are rational.

Some economists subscribe to the *efficient markets hypothesis*, according to which the market price of a company's stock is the rational valuation of the company's value, given current information about the company's business prospects. This hypothesis rests on two foundations:

1. Each company listed on a major stock exchange is followed closely by many professional portfolio managers. Every day, these managers monitor news stories to judge the company's value. Their job is to buy a stock when its price falls below its value and to sell it when its price rises above its value.
2. A stock's price is set by the equilibrium of supply and demand. At the market price, the number of shares being offered for sale equals the number of shares that people want to buy. That is, at the market price, the number of people who think the stock is overvalued balances the number of people who think it's undervalued. As judged by the typical person in the market, the stock must be fairly valued.

According to this theory, the stock market is *informationally efficient*: it reflects all available information about the asset's value. Stock prices change when information changes. When there is good news about a company's prospects, its stock price rises. When the company's prospects deteriorate, the price falls. But at any moment, the market price is the best guess of the company's value.

An implication of this hypothesis is that stock prices should follow a *random walk*, meaning that changes in stock prices should be impossible to predict. If a person could reliably predict using publicly available information that a stock price would rise by 10 percent tomorrow, the stock market would be failing to incorporate that information today. The only thing that should move stock prices is news that changes the market's perception of the company's value. But such news must be unpredictable; otherwise, it wouldn't really be news. Thus, changes in stock prices should be unpredictable.

What is the evidence for the efficient markets hypothesis? Its proponents note that it is hard to beat the market. Statistical tests show that stock prices are random walks—or at least approximately so. Moreover, index funds (which buy stocks from all companies in a stock market index) outperform most actively managed mutual funds run by professional money managers (who try to buy stocks selling below their true value).

Yet many economists are skeptical that the stock market is rational. The skeptics point out that many changes in stock prices are hard to attribute to news. They suggest that when buying and selling, stock investors are focused less on companies' fundamental values and more on what they expect other investors will later pay.

John Maynard Keynes proposed a famous analogy to explain this speculation. In his day, some newspapers held "beauty contests" in which they printed the pictures of 100 women, and readers were invited to submit a list of the five most beautiful. A prize went to the reader whose choices most closely matched those of the consensus of the other entrants. Naive entrants would have simply picked those they considered the most beautiful. But a slightly more sophisticated strategy would have been to guess the five women whom other people considered the most beautiful. Other people, however, were likely thinking along the same lines. So an even more sophisticated strategy would have been to guess who other people thought other people thought were the most beautiful. And so on. In the end, judging true beauty was less important to winning the contest than guessing other people's opinions about other people's opinions.

Similarly, Keynes reasoned that because stock investors will eventually sell their shares to others, they are more concerned about other people's valuation of a company than about its true worth. The best investors, in his view, are those who are good at outguessing mass psychology. He believed that changes in stock prices often reflect irrational waves of optimism and pessimism, which he called the *animal spirits* of investors.

These two views of the stock market persist to this day. Some economists see the stock market through the lens of the efficient markets hypothesis. Others believe that irrational speculation is the norm. In their view, the stock market often moves for no good reason, and because the stock market affects aggregate demand, changes in stock prices are a source of short-run fluctuations in output and employment.[2]

In 2008 and 2009, the crucial asset was residential real estate. The average price of houses in the United States had experienced a boom earlier in the decade. This boom was driven in part by lax lending standards; many *subprime* borrowers—those with particularly risky credit profiles—were lent money to buy a house while offering only a very small down payment. In essence, the financial system failed to do its job of dealing with asymmetric information by making loans to many borrowers who, it turned out, would later have trouble making their mortgage payments. The housing boom was also encouraged by government policies promoting homeownership and was fed by excessive optimism on the part of home-buyers, who thought prices would rise forever. The housing boom, however, proved unsustainable. Over time, the number of homeowners falling behind on their mortgage payments rose, and sentiment among home-buyers shifted. House prices fell by about 30 percent from 2006 to 2009. The nation had not experienced such a large decline in house prices since the 1930s.

2. Insolvencies at Financial Institutions

A large decline in asset prices may cause problems at banks and other financial institutions. To ensure that borrowers repay their loans, banks often require them to post collateral. That is, a borrower has to pledge assets that the bank can seize if the borrower defaults. Yet when assets decline in price, the collateral falls in value, perhaps below the amount of the loan. In this case, if the borrower defaults on the loan, the bank may be unable to recover its money.

As we discussed in Chapter 4, banks rely heavily on **leverage**, the use of borrowed funds for the purposes of investment. Leverage amplifies the positive and negative effects of asset returns on a bank's financial position. A key number is the *leverage ratio*: the ratio of bank assets to bank capital. A leverage ratio of 20, for example, means that for every $1 in capital put into the bank by its owners, the bank has borrowed (via deposits and other loans) $19, allowing the bank to hold $20 in assets. In this case, if defaults cause the value of the bank's assets to fall by 2 percent, then the bank's capital will fall by 40 percent. If the value of bank assets falls by more than 5 percent, then its assets will fall below its liabilities, and the bank will not have the resources to repay all its depositors and other creditors. The bank is said to be *insolvent*. Widespread insolvency within the financial system is the second element of a financial crisis.

In 2008 and 2009, many banks and other financial firms had in effect placed bets on real estate prices by holding mortgages backed by that real estate. They assumed that house prices would keep rising or at least hold steady, so the collateral backing these loans would ensure their repayment. When house prices fell, however, large numbers of homeowners found themselves *underwater*: the value of their homes was less than the amount they owed on their mortgages. When homeowners stopped paying their mortgages, the banks could foreclose on the houses, but they could recover only a fraction of what they were owed. These defaults pushed several financial institutions toward bankruptcy. These institutions included major investment banks (Bear Stearns and Lehman Brothers), government-sponsored enterprises involved in the mortgage market (Fannie Mae and Freddie Mac), and a large insurance company (AIG).

3. Falling Confidence

The third element of a financial crisis is a decline in confidence in financial institutions. While some deposits in banks are insured by government policies, not all are. As insolvencies mount, every financial institution becomes a candidate for the next bankruptcy. People with uninsured deposits in those institutions pull out their money. Facing a rash of withdrawals, banks reduce lending and start selling off some of their assets to increase their cash reserves.

As banks sell assets, they depress the market prices of those assets. Because buyers of risky assets are scarce in the midst of a crisis, the assets' prices can fall precipitously. This phenomenon is called a **fire sale**, similar to the reduced prices that a store might charge to get rid of merchandise quickly after a fire. These fire-sale prices, however, cause problems at other banks. Accountants and regulators may require these banks to revise their balance sheets and reduce the reported value of their own holdings of these assets. In this way, problems in one bank can spread to others.

In 2008 and 2009, the financial system was seized by great uncertainty about where the insolvencies would stop. The collapse of the giants Bear Stearns and Lehman Brothers made people wonder whether other large financial firms, such as Morgan Stanley, Goldman Sachs, and Citigroup, would meet a similar fate. The problem was exacerbated by the firms' interdependence. Because they had many contracts with one another, the demise of any one of these institutions would undermine the others. Moreover, because of the complexity of the arrangements, depositors could not be sure how vulnerable these firms were. The lack of transparency fed the crisis of confidence.

FYI

The TED Spread

A common type of indicator of perceived credit risk is the difference between two interest rates of similar maturity. For example, Financially Shaky Corporation might have to pay 7 percent for a one-year loan, whereas Safe and Solid Corporation has to pay only 3 percent. That spread of 4 percentage points occurs because lenders are worried that Financially Shaky might default; as a result, they demand compensation for bearing that risk. If Financially Shaky gets some bad news about its financial position, the interest rate spread might rise to 5 or 6 percentage points or even higher. Thus, one way to monitor perceptions of credit risk is to follow interest rate spreads.

One noteworthy interest rate spread is the TED spread (and not just because it rhymes). The TED spread is the difference between three-month interbank loans and three-month Treasury bills. The T in TED stands for T-bills, and ED stands for EuroDollars (because, for regulatory reasons, these interbank loans typically take place in London). The TED spread is measured in basis points, where a basis point is one one-hundredth of a percentage point (0.01 percent). Normally, the TED spread is about 20 to 40 basis points (0.2 to 0.4 percent). The spread is small because commercial banks, while a bit riskier than the government, are still very safe. Lenders do not require much extra compensation to accept the debt of banks rather than that of the government.

During a financial crisis, however, confidence in the banking system falls. As a result, banks become reluctant to lend to one another, and the TED spread rises substantially. Figure 18-1 shows the TED spread before, during, and after the financial crisis of 2008–2009. As the crisis unfolded, the TED spread rose substantially, reaching 458 basis points in October 2008, shortly after the investment bank Lehman Brothers declared bankruptcy. The high level of the TED spread is a direct indicator of how worried people were about the solvency of the banking system.

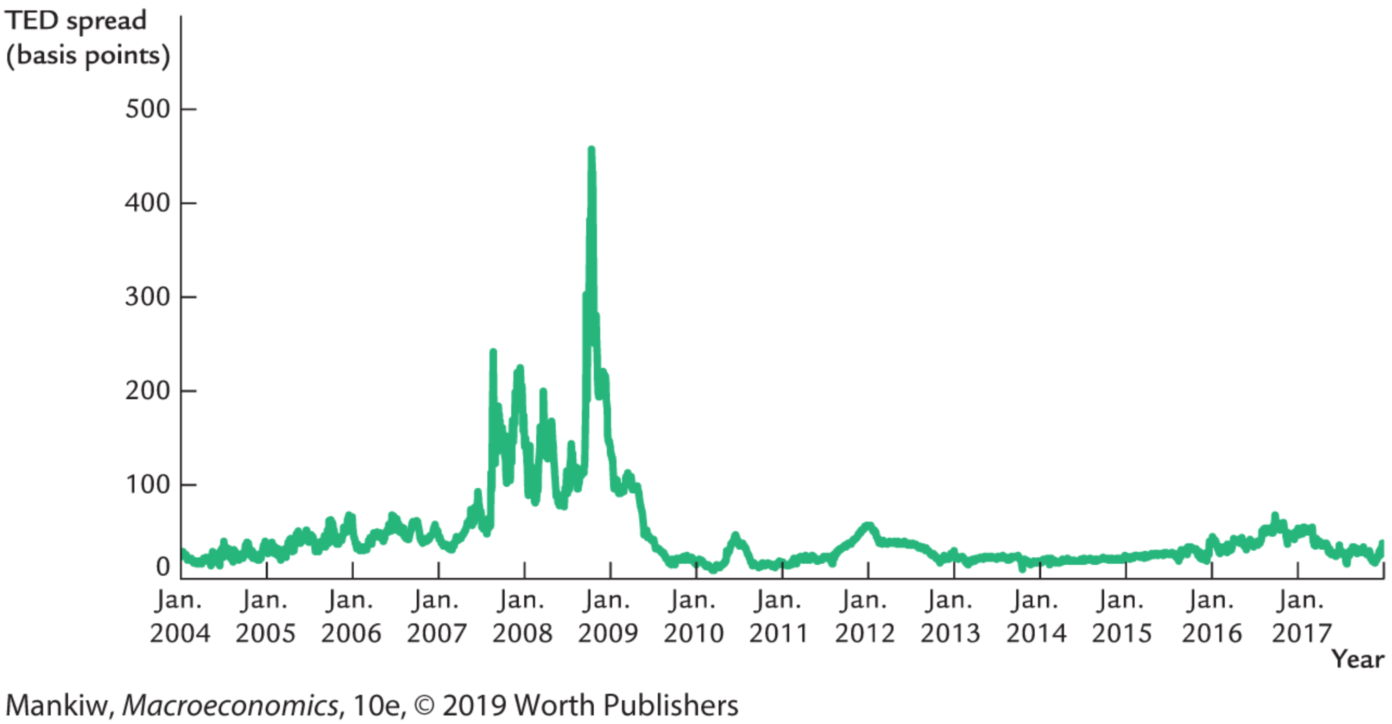


FIGURE 18-1 The TED Spread The TED spread is the difference between the interest rate on three-month interbank loans and the interest rate on three-month Treasury bills. It rises when lending to banks is considered particularly risky.

Data from: Federal Reserve Bank of St. Louis.

4. Credit Crunch

The fourth element of a financial crisis is a credit crunch. With many financial institutions facing difficulties, would-be borrowers have trouble getting loans, even if they have profitable investment projects. In essence, the financial system has trouble performing its normal function of directing the resources of savers into the hands of borrowers with the best investment opportunities.

The tightening of credit was clear during the 2008–2009 financial crisis. Not surprisingly, as banks realized that house prices were falling and that previous lending standards had been too lax, they started raising standards for those applying for mortgages. They required larger down payments and scrutinized borrowers' financial information more closely. But the reduction in lending did not affect only home-buyers. Small businesses found it harder to borrow to finance business expansions or to buy inventories. Consumers found it harder to qualify for credit cards or car loans. Thus, banks responded to their own financial problems by becoming more cautious in all kinds of lending.

5. Recession

The fifth element of a financial crisis is an economic downturn. With people unable to obtain consumer credit and firms unable to obtain financing for new investment projects, the demand for goods and services declines. Within the context of the *IS–LM* model, this event can be interpreted as contractionary shifts in the consumption and investment functions, which cause similar shifts in the *IS* curve and the aggregate demand curve. As a result, national income falls, and unemployment rises.

Indeed, the Great Recession of 2008–2009 was a deep one. Unemployment rose from about 4.5 percent in early 2007 to 10 percent in late 2009. Worse yet, it lingered at a high level for a long time. Even after the recovery officially began in June 2009, growth in GDP was so meager that unemployment declined only slightly. The unemployment rate remained above 8 percent until late 2012.

6. A Vicious Circle

The sixth and final element of a financial crisis is a vicious circle. The economic downturn reduces the profitability of many companies and the value of many assets. The stock market declines. Some firms go bankrupt and default on their business loans. Many workers become unemployed and default on their personal loans. Thus, we return to steps 1 (asset-price busts) and 2 (financial institution insolvencies). The problems in the financial system and the economic downturn reinforce each other. Figure 18-2 illustrates the process.

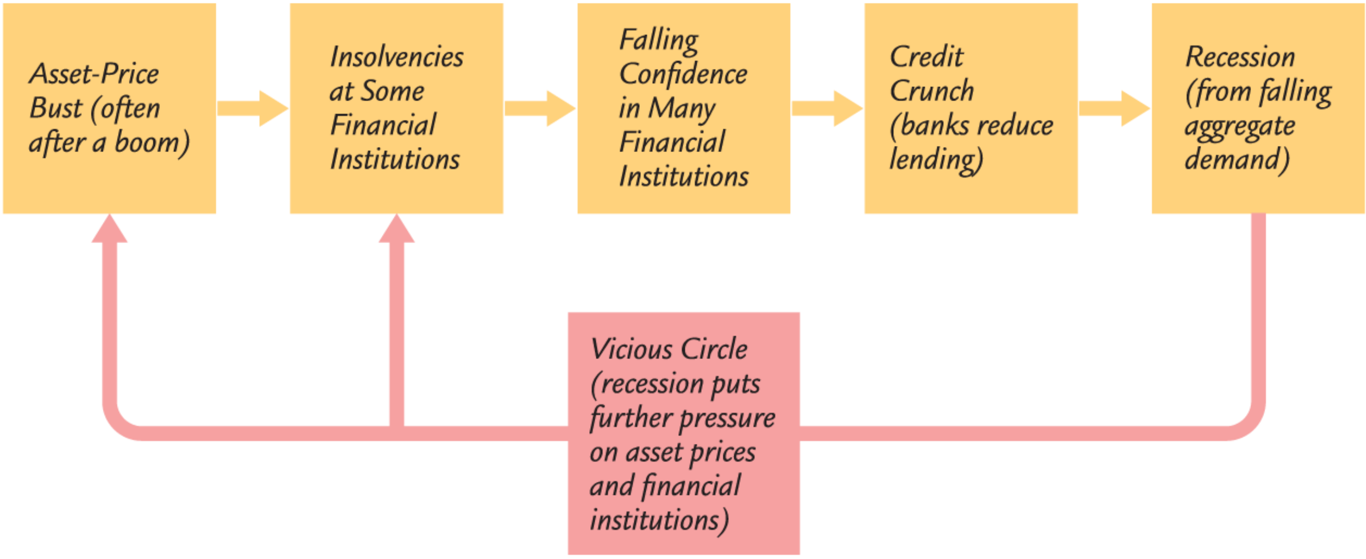


FIGURE 18-2 The Anatomy of a Financial Crisis This figure is a schematic illustration of the six elements of a financial crisis.

In 2008 and 2009, the vicious circle was apparent. Some feared that the combination of a weakening financial system and a weakening economy would cause the economy to spiral out of control, leading to another Great Depression. Fortunately, that did not occur, in part because policymakers were intent on preventing it.

That brings us to the next question: Faced with a financial crisis, what can policymakers do?

CASE STUDY

Who Should Be Blamed for the Financial Crisis of 2008–2009?

"Victory has a hundred fathers, and defeat is an orphan." This famous quotation from John F. Kennedy contains a perennial truth. Everyone is eager to take credit for success, but no one wants to accept blame for failure. In the aftermath of the financial crisis of 2008–2009, many people wondered who was to blame. No one stepped forward to take responsibility.

Nonetheless, observers have pointed their fingers at many possible culprits. The accused include the following:

- *The Federal Reserve*. The nation's central bank kept interest rates low in the aftermath of the 2001 recession. While this policy helped promote the recovery, it also encouraged households to borrow and buy housing. Some economists believe that, by keeping interest rates too low for too long, the Fed contributed to the housing bubble that led to the financial crisis.
- *Home-buyers*. Many people were reckless in borrowing more than they could afford to repay. Others bought houses as a gamble, hoping that house prices would continue their rapid increase. When house prices fell instead, many of these homeowners defaulted on their debts.
- *Mortgage brokers*. Many providers of home loans encouraged households to borrow excessively. Sometimes they pushed complicated mortgage products with payments that were low initially but exploded later. Some offered what were called NINJA loans (an acronym for "no income, no job or assets") to households that should not have qualified for a mortgage. The brokers did not hold these risky loans but

instead sold them for a fee after they were issued.

- *Investment banks*. Many of these financial institutions packaged bundles of risky mortgages into mortgage-backed securities and then sold them to buyers (such as pension funds) that were not fully aware of the risks they were taking on.
- *Rating agencies*. The agencies that evaluated the riskiness of debt instruments gave high ratings to various mortgage-backed securities that later turned out to be highly risky. With the benefit of hindsight, it is clear that the models the agencies used to evaluate the risks were based on dubious assumptions.
- *Regulators*. Regulators of banks and other financial institutions are supposed to ensure that these firms do not take undue risks. Yet the regulators failed to appreciate that a large decline in house prices might occur and that, if it did, it could have implications for the entire financial system.
- *Government policymakers*. For many years, political leaders have pursued policies promoting homeownership, including the tax deductibility of mortgage interest and the establishment of Fannie Mae and Freddie Mac (the government-sponsored enterprises that supported mortgage lending). Households with shaky finances, however, might have been better off renting.

In the end, it seems that each of these groups (and perhaps a few others as well) bear some of the blame. As *The Economist* magazine once put it, the problem was one of "layered irresponsibility."

Finally, remember that this financial crisis was not the first one in history. Such events, though fortunately rare, occur from time to time. Rather than look for a culprit to blame for this singular event, perhaps we should view speculative excess and its ramifications as an inherent feature of market economies. Policymakers can respond to financial crises as they happen, and they can take steps to reduce the likelihood and severity of such crises, but preventing them entirely may be too much to ask, given our current knowledge.[3] ■

Policy Responses to a Crisis

Because financial crises are both severe and multifaceted, economic policymakers use various tools, often simultaneously, to limit the damage. Here we discuss three broad categories of policy responses.

Conventional Monetary and Fiscal Policy

As we have seen, financial crises raise unemployment and lower incomes because they lead to a contraction in the aggregate demand for goods and services. Policymakers can mitigate these effects by using the tools of monetary and fiscal policy to expand aggregate demand. The central bank can increase the money supply and lower interest rates. The government can increase government spending and cut taxes. That is, a financial crisis can be seen as a shock to the aggregate demand curve, which can, to some degree, be offset by appropriate monetary and fiscal policy.

Policymakers did precisely this during the financial crisis of 2008–2009. To expand aggregate demand, the Fed cut its target for the federal funds rate from 5.25 percent in September 2007 to approximately zero in December 2008. The federal funds rate then stayed at that low level for the next six years. In February 2008

President Bush signed into law a $168 billion stimulus package, providing tax rebates of $300 to $1,200 for every taxpayer. In 2009 President Obama signed into law a $787 billion stimulus, which included both tax reductions and increases in government spending. All of these moves were aimed at propping up aggregate demand.

There are limits, however, to how much conventional monetary and fiscal policy can do. A central bank cannot cut its interest rate target much below zero. (Recall the discussion of the *liquidity trap* in Chapter 12.) Fiscal policy is limited as well. Stimulus packages add to the government budget deficit, which is already enlarged because economic downturns automatically increase unemployment-insurance payments and decrease tax revenue. Increases in government debt are a concern because they place a burden on future generations of taxpayers and call into question the government's own solvency. In the aftermath of the financial crisis of 2008–2009, the federal government's budget deficit reached levels not seen since World War II. Recall from Chapter 17 that in August 2011 Standard & Poor's responded to the fiscal imbalance by reducing its rating on U.S. government debt below the top AAA level for the first time in the nation's history, a decision that made additional fiscal stimulus more difficult.

The limits of monetary and fiscal policy during a financial crisis can lead policymakers to consider other, and sometimes unusual, alternatives. These other types of policy are of a different nature. Rather than address the symptom of a financial crisis (a decline in aggregate demand), they aim to fix the financial system itself. If the normal process of financial intermediation can be restored, consumers and businesses will be able to borrow again, and the economy's aggregate demand will recover. The economy can then return to full employment and rising incomes. The next two categories describe the major policies aimed at fixing the financial system.

Lender of Last Resort

When people lose confidence in a bank, they withdraw their deposits. In a system of fractional-reserve banking, large and sudden withdrawals can be a problem. Even if the bank is solvent (meaning that the value of its assets exceeds the value of its liabilities), it may have trouble satisfying all its depositors' requests. Many of the bank's assets are illiquid—that is, they cannot be easily sold and turned into cash. A business loan to a local restaurant, a car loan to a local family, and a student loan to your roommate, for example, may be valuable assets to the bank, but they cannot be easily used to satisfy depositors who are demanding their money back immediately. A situation in which a solvent bank has insufficient funds to satisfy its depositors' withdrawals is called a **liquidity crisis**.

The central bank can remedy this problem by lending money directly to the bank. As we discussed in Chapter 4, the central bank can create money out of thin air by, in effect, printing it. (Or, more realistically in our electronic era, it creates a bookkeeping entry for itself that represents those monetary units.) It can then lend this newly created money to the bank experiencing greater-than-normal withdrawals and accept the bank's illiquid assets as collateral. When a central bank lends to a bank in the midst of a liquidity crisis, it is said to act as a **lender of last resort**.

The goal of such a policy is to allow a bank experiencing high withdrawals to weather the storm of reduced confidence. Without such a loan, the bank might be forced to sell its illiquid assets at fire-sale prices. If such a fire sale were to occur, the value of the bank's assets would decline, and a liquidity crisis could then threaten the bank's solvency. By acting as a lender of last resort, the central bank stems the problem of bank insolvency and helps restore the public's confidence in the banking system.

During 2008 and 2009, the Fed was very active as a lender of last resort. As we discussed in Chapter 4, such activity traditionally takes place at the Fed's discount window, through which the Fed lends to banks at its discount rate. During this crisis, however, the Fed set up a variety of new ways to lend to financial institutions. The financial institutions included were not only traditional commercial banks but also shadow banks. **Shadow banks** are a diverse set of financial institutions that perform some functions similar to those of banks but do so outside the regulatory system that applies to traditional banking. Because the shadow banks were experiencing difficulties similar to those of commercial banks, the Fed was concerned about these institutions as well.

For example, from October 2008 to October 2009, the Fed was willing to make loans to money market mutual funds. Money market funds are not banks, and they do not offer insured deposits. But they are in some ways similar to banks: they take in deposits, invest the proceeds in short-term loans such as commercial paper issued by corporations, and assure depositors that they can obtain their deposits on demand with interest. In the midst of the financial crisis, depositors worried about the value of the assets the money market funds had purchased, so these funds experienced substantial withdrawals. The shrinking deposits in money market funds meant that there were fewer buyers of commercial paper, making it hard for firms that needed the proceeds from these loans to finance their business operations. By its willingness to lend to money market funds, the Fed helped maintain this form of financial intermediation.

It is not crucial to learn the details of all the lending facilities the Fed established during the crisis. Many of these programs were ended as the economy recovered because they were no longer needed. What is important to understand is that these programs, both old and new, had one purpose: to ensure that the financial system remained liquid. That is, as long as a financial institution had assets that could serve as reliable collateral, the Fed stood ready to lend it money so that its depositors could withdraw their funds.

Injections of Government Funds

The final category of policy responses to a financial crisis involves the government's use of public funds to prop up the financial system.

The most direct action of this sort is a giveaway of public funds to those who have experienced losses. Deposit insurance is an example. Through the Federal Deposit Insurance Corporation (FDIC), the federal government promises to make up for losses that a depositor experiences when a bank becomes insolvent. In 2008, the FDIC increased the maximum deposit it would cover from $100,000 to $250,000 to reassure bank depositors that their funds were safe.

Giveaways of public funds can also occur on a more discretionary basis. For example, in 1984 a large bank called Continental Illinois found itself on the brink of insolvency. Because Continental Illinois had many relationships with other banks, regulators feared that allowing it to fail would threaten the entire financial system. As a result, the FDIC promised to protect all of its depositors, not just those under the insurance limit. Eventually, it bought the bank from shareholders, added capital, and sold it to Bank of America. This policy operation cost taxpayers about $1 billion. It was during this episode that a member of Congress coined the phrase "too big to fail" to describe a firm so central to the economy that policymakers would not allow it to enter bankruptcy.

Another way for the government to inject public funds is to make risky loans. Normally, when the Fed acts as lender of last resort, it does so by lending to a financial institution that can pledge good collateral. But if the government makes loans that might not be repaid, it is putting public funds at risk. If the borrowers default, taxpayers end up losing.

During the financial crisis of 2008–2009, the Fed engaged in a variety of risky lending. In March 2008, it made a $29 billion loan to JPMorgan Chase to facilitate its purchase of the nearly insolvent Bear Stearns. The only collateral the Fed received was Bear's holdings of mortgage-backed securities, which were of dubious value. Similarly, in September 2008, the Fed loaned $85 billion to prop up the insurance giant AIG, which faced large losses from having insured the value of some mortgage-backed securities (through an agreement called a *credit default swap*). The Fed took these actions to prevent Bear Stearns and AIG from entering a long bankruptcy process, which could have further threatened the financial system.

A final way for the government to use public funds to address a financial crisis is for the government itself to inject capital into financial institutions. In this case, rather than being just a creditor, the government gets an ownership stake in the companies. The AIG loans in 2008 had significant elements of this: as part of the loan deal, the government got warrants (options to buy stock) and eventually owned most of the company. (The shares were sold several years later.) Another example is the capital injections organized by the U.S. Treasury in 2008 and 2009. As part of the Troubled Asset Relief Program (TARP), the government put hundreds of billions of dollars into various banks in exchange for equity shares in those banks. The goal of the program was to maintain the banks' solvency and protect the process of financial intermediation.

Not surprisingly, the use of public funds to prop up the financial system, whether done with giveaways, risky lending, or capital injections, is controversial. Critics argue that it is unfair to taxpayers to use their resources to rescue financial market participants from their own mistakes. Moreover, the prospect of bailouts may increase moral hazard because when people believe the government will cover their losses, they are more likely to take excessive risks. Financial risk taking becomes “heads I win, tails the taxpayers lose.” Advocates of these policies acknowledge these problems, but they point out that risky lending and capital injections could make money for taxpayers if the economy recovers. (Indeed, in December 2014, the federal government estimated that TARP yielded a $15 billion profit.) More important, they believe that the costs of these policies are more than offset by the benefits of averting a deeper crisis and more severe economic downturn.

Policies to Prevent Crises

In addition to the question of how policymakers should respond when facing a financial crisis, there is another key policy debate: how should policymakers prevent future financial crises? Unfortunately, there is no easy answer. But here are five areas where policymakers have been considering their options and, in some cases, revising their policies.

Focusing on Shadow Banks

Traditional commercial banks are heavily regulated. One justification is that the government insures some of their deposits through the FDIC. Policymakers have long understood that deposit insurance creates a moral hazard problem. Because of deposit insurance, depositors have no incentive to monitor the riskiness of banks in which they make their deposits; as a result, bankers have an incentive to make excessively risky loans, knowing they will reap any gains, and the deposit insurance system will cover any losses. In response to this moral hazard problem, the government regulates the risks that banks take.

Much of the crisis of 2008–2009, however, concerned not traditional banks but rather *shadow banks*—financial institutions that (like banks) are at the center of financial intermediation but (unlike banks) do not take in deposits insured by the FDIC. Bear Sterns and Lehman Brothers, for example, were investment banks and, therefore, subject to less regulation. Similarly, hedge funds, insurance companies, and private equity firms can be considered shadow banks. These institutions do not suffer from the traditional problem of moral hazard arising from deposit insurance, but the risks they take may nonetheless be a concern of public policy because their failure can have macroeconomic ramifications.

Many policymakers have suggested that these shadow banks should be limited in how much risk they take. One way to do that would be to require that they hold more capital, which would reduce these firms’ ability to use leverage. Advocates of this idea say it would enhance financial stability. Critics say it would restrict these institutions' ability to do their job of financial intermediation.

Another issue concerns what happens when a shadow bank runs into trouble and nears insolvency. The Dodd-Frank Act, passed in 2010, gave the FDIC *resolution authority* over shadow banks, much as it already had over traditional commercial banks. That is, the FDIC can now take over and close a nonbank financial institution if it is concerned that the institution is having trouble and could create systemic risk for the economy. Advocates of this new law believe it will establish a more orderly process when a shadow bank fails and thereby prevent a more general loss of confidence in the financial system. Critics fear it will make bailouts of these institutions with taxpayer funds more common and thus exacerbate moral hazard.

Restricting Size

The financial crisis of 2008–2009 centered on a few large financial institutions. Some economists believe the problem would have been averted, or at least less severe, if the financial system had been less concentrated. When a small institution fails, bankruptcy law can take over as it usually does, adjudicating the claims of the various stakeholders, without resulting in economy-wide problems. These economists argue that if a financial institution is too big to fail, it is too big.

Various ideas have been proposed for limiting the size of financial firms. One would be to restrict mergers among banks. (Over the past half century, the banking industry has become vastly more concentrated, largely through bank mergers.) Another idea is to impose higher capital requirements on larger banks. Advocates of these ideas say that a financial system with smaller firms would be more stable. Critics say that such a policy would prevent banks from taking advantage of economies of scale and that the higher costs would be passed on to the banks' customers.

Reducing Excessive Risk Taking

The financial firms that failed during the financial crisis of 2008–2009 did so because they took risks that resulted in large losses. Some observers believe that one way to reduce the risk of future crises is to limit excessive risk taking. Yet because risk taking is at the heart of what many financial institutions do, drawing the line between appropriate and excessive risks is not easy.

Nonetheless, the Dodd-Frank Act included several provisions aimed at limiting risk taking. Perhaps the best known is the Volcker rule, named after Paul Volcker, the former Fed chair who first proposed it. Under the Volcker rule, commercial banks are restricted from making certain kinds of speculative investments. Advocates say the rule will help protect banks. Critics say that by restricting the banks' trading activities, it will make the market for those speculative financial instruments less liquid.

In addition, the bank regulators at the Fed now require that large banks undergo regular *stress tests*. To conduct these tests, the regulators posit a hypothetical scenario of economic distress, such as a rise in unemployment to 10 percent, a 20 percent drop in the stock market, and a 30 percent plunge in house prices. Each bank is then asked to estimate what would happen to the value of its assets in this scenario. The goal is to make sure that the bank has sufficient capital to weather the storm. If it doesn't, the bank must either raise more capital or reduce the riskiness of its assets. These stress tests are one gauge of whether a bank has taken on excessive risks, but because they are based on hypothetical scenarios, their value is limited by the regulators' ability to imagine the adverse outcomes that might occur.

Making Regulation Work Better

The financial system is diverse, with many firms performing various functions and having developed at different stages of history. As a result, the regulatory apparatus overseeing these firms is fragmented. The Fed, the Office of the Comptroller of the Currency, and the FDIC all regulate commercial banks. The Securities and Exchange Commission regulates investment banks and mutual funds. Individual state agencies regulate insurance companies.

After the financial crisis of 2008–2009, policymakers tried to improve the system of regulation. The Dodd-Frank Act created the new Financial Services Oversight Council, chaired by the Secretary of the Treasury, to coordinate the regulatory agencies. It also created the new Office of Credit Ratings to oversee the private credit rating agencies, which were blamed for the failure to anticipate the risk in many mortgage-backed securities. The law also established the new Consumer Financial Protection Bureau, with the goal of ensuring fairness and transparency in how financial firms market their products to consumers. Because financial crises are infrequent events, often occurring decades apart, it will take a long time to tell whether this new regulatory structure works better than the old one.

Taking a Macro View of Regulation

Policymakers have increasingly taken the view that the regulation of financial institutions requires more of a macroeconomic perspective. Traditionally, financial regulation has been **microprudential**: its goal has been to reduce the risk of distress in individual financial institutions, thereby protecting the depositors and other stakeholders in those institutions. Today, financial regulation is also **macroprudential**: its goal is also to reduce the risk of system-wide distress, thereby protecting the overall economy against declines in production and employment. Microprudential regulation takes a bottom-up approach by focusing on individual institutions and assessing the risks that each of them faces. By contrast, macroprudential regulation takes a top-down approach by focusing on the big picture and assessing the risks that can affect many financial institutions at the same time.

For example, macroprudential regulation could have addressed the boom and bust in the housing market that was the catalyst for the 2008–2009 financial crisis. Advocates of such regulation argue that as house prices increased, policymakers should have required a larger down payment when a home-buyer purchased a house with a mortgage. This policy might have slowed the speculative bubble in house prices, and it would have led to fewer mortgage defaults when house prices later declined. Fewer mortgage defaults, in turn, would have helped protect many financial institutions that had acquired stakes in housing-related securities. Critics of such a policy question whether government regulators are sufficiently adept to identify and remedy economy-wide risks. They worry that attempts to do so would add to the regulatory burden; an increase in required down payments, for instance, makes it harder for less wealthy families to buy their own homes.

Without doubt, in light of what was learned during and after the financial crisis, financial regulators will pay renewed attention to macroeconomic stability as one of their goals. In this sense, macroprudential regulation takes its place alongside the traditional tools of monetary and fiscal policy. How active policymakers should be in using this tool, however, remains open to debate.[4]

CASE STUDY

The European Sovereign Debt Crisis

As the United States was beginning to recover from its financial crisis of 2008–2009, another crisis erupted in the eurozone, the part of Europe that uses the euro as a common currency. The problem stemmed from debt issued by governments, called *sovereign debt*. For many years, banks and bank regulators had treated such debt as risk-free. They presumed that the central governments of Europe would always honor their obligations. Because of this belief, these bonds paid a lower interest rate and commanded a higher price than they would have if they had been perceived as less reliable credit risks.

In 2010, however, financial market participants started to doubt this optimism about European governments. The problem began with Greece. The debt (net financial liabilities) of the Greek government had increased to 116 percent of its GDP in 2010, twice the European average. Moreover, it became apparent that for years Greece had been misreporting its finances and that it had no plan to rein in its soaring debts. In April 2010, Standard & Poor’s reduced the rating on Greek debt to junk status, indicating a poor credit risk. Because many feared that default was likely, the prices of Greek debt fell, and the interest rate that Greece had to pay on new borrowing rose markedly. By the summer of 2011, the interest rate on Greek debt was 26 percent. In November of that year, it rose to over 100 percent.

European policymakers were concerned that problems in Greece could have repercussions throughout Europe. Many European banks held Greek debt among their assets. As the value of Greek debt fell, the banks were pushed toward insolvency. A Greek default could send many banks over the edge, leading to a broader crisis of confidence. As a result, policymakers in healthier European economies, such as Germany and France, helped arrange continuing loans to Greece to prevent immediate default. Some of these loans were from the European Central Bank, which controls monetary policy in the eurozone.

This policy move was unpopular. Voters in Germany and France wondered why their tax dollars should help rescue the Greeks from their own profligacy. Voters in Greece, meanwhile, were angry because these loans came with the conditions that Greece drastically cut government spending and raise taxes. These austerity measures led to rioting in Greek streets.

Making matters worse, Greece was not the only country with such problems. If Greece was allowed to default, rather than being bailed out by its richer neighbors, some feared that Portugal, Ireland, Spain, and Italy might be close behind. A widespread decline in the value of the sovereign debt of all these nations would put serious strains on the European banking system. And since the world's banking systems are interconnected, it would also put strains on the rest of the world.

The policy actions in response to this crisis were successful in one sense: despite predictions that Greece and other problematic countries might stop using the euro as their currency, the monetary union was maintained. But the pain resulting from the crisis was nonetheless substantial and long-lasting. In 2013, the unemployment rate was 27 percent in Greece, 26 percent in Spain, and 16 percent in Portugal (but only 5 percent in Germany, the most populous eurozone nation). As a standard Phillips curve predicts, the economic slack pulled inflation in Europe well below the target rate of 2 percent. From 2014 to 2016, inflation in the eurozone was only slightly above zero. To expand aggregate demand and stimulate the economy, the European Central Bank cut the interest rate to about zero as the crisis unfolded. In addition, in early 2015, the ECB announced a program of quantitative easing, under which it would buy large quantities of government bonds to reduce longer-term interest rates and further expand aggregate demand.

As this book was going to press in early 2018, the end of the story was yet to be written, but preliminary evidence suggested that the ECB's policies were working. Inflation in the eurozone was still below target but was closer to 2 percent. Unemployment rates remained high, at 21 percent in Greece, 16 percent in Spain, and 8 percent in Portugal, but they were below the levels reached five years earlier. In short, the European economies appeared to be heading in the right direction.[5] ■

18-3 Conclusion

Throughout history, financial crises have been a major source of economic fluctuations and a main driver of economic policy. In 1873 Walter Bagehot published a celebrated book called *Lombard Street* about how the Bank of England should manage financial crises. His recommendation that it should act as a lender of last resort has over time become the conventional wisdom. In 1913, in the aftermath of the banking panic of 1907, Congress passed the act establishing the Federal Reserve. Congress wanted the new central bank to oversee the banking system to ensure greater financial and macroeconomic stability.

The Fed has not always succeeded in achieving this goal. Many economists believe that the Great Depression of the 1930s was so severe because the Fed failed to follow Bagehot's advice. If it had been a more active lender of last resort, the crisis of confidence in the banks and the resulting collapse in the money supply and aggregate demand might have been averted. Mindful of this history, the Fed was more active in trying to mitigate the impact of the financial crisis of 2008–2009.

After a crisis, it is easy to lament the problems caused by the financial system, but we should remember the benefits that the system brings. The financial system gives savers the ability to earn the best possible return at the lowest possible risk. It gives entrepreneurs the ability to fund new business ventures. By bringing together those who want to save and those who want to invest, the financial system promotes economic growth and overall prosperity.

CHAPTER 19

The Microfoundations of Consumption and Investment

> Consumption is the sole end and purpose of all production.
>
> —Adam Smith

> The social object of skilled investment should be to defeat the dark forces of time and ignorance which envelope our future.
>
> —John Maynard Keynes

How do households decide how much of their income to consume today and how much to save for the future? How do firms decide how much to invest to expand their stock of capital? These are microeconomic questions because they address the behavior of individual decisionmakers. Yet their answers have macroeconomic consequences. As we have seen in previous chapters, households' consumption decisions and firms' investment decisions affect the behavior of the economy as a whole.

In previous chapters we explained consumption and investment with simple functions: $C = C(Y - T)$ and $I = I(r)$. These showed that consumption depends on disposable income and that investment depends on the real interest rate, and they allowed us to develop models for long-run and short-run analysis. But they are too simple to fully explain consumer and firm behavior. In this chapter we examine the consumption and investment functions in greater detail and develop a more thorough explanation of what determines spending by households and firms.

As we discussed in Chapter 1, the field of economics is divided into two broad subfields—microeconomics and macroeconomics. Yet sometimes it is best to break down the wall that separates these subfields. In this chapter we see how studying the microeconomic foundations of consumption and investment decisions can enhance our understanding of macroeconomic events and policy.

19-1 What Determines Consumer Spending?

Since macroeconomics began as a field of study, many economists have proposed ways of explaining consumer behavior. Here we present the views of five prominent economists.

John Maynard Keynes and the Consumption Function

We begin with John Maynard Keynes's *General Theory*, published in 1936. Keynes made the consumption function central to his theory of economic fluctuations, and it has played a key role in macroeconomic analysis ever since. Let's consider what Keynes thought about the consumption function and then see what puzzles arose when his ideas were confronted with the data.

Keynes's Conjectures

Today, economists who study consumption rely on sophisticated techniques of data analysis. With the help of computers, they analyze aggregate data on the behavior of the overall economy from the national income accounts and detailed data on the behavior of individual households from surveys. Because Keynes wrote in the 1930s, however, he had neither the advantage of these data nor the computers necessary to analyze such large data sets. Instead of relying on statistical analysis, Keynes made conjectures about the consumption function based on introspection and casual observation.

First and most important, Keynes conjectured that the **marginal propensity to consume**—the amount consumed out of an additional dollar of income—is between zero and one. He wrote that the "fundamental psychological law, upon which we are entitled to depend with great confidence, . . . is that men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income." That is, when a person earns an extra dollar, she typically spends some of it and saves some of it. As we saw in Chapter 11 when we developed the Keynesian cross, the marginal propensity to consume was crucial to Keynes's advice for how to reduce widespread unemployment. The power of fiscal policy to influence the economy—as expressed by the fiscal-policy multipliers—arises from the feedback between income and consumption.

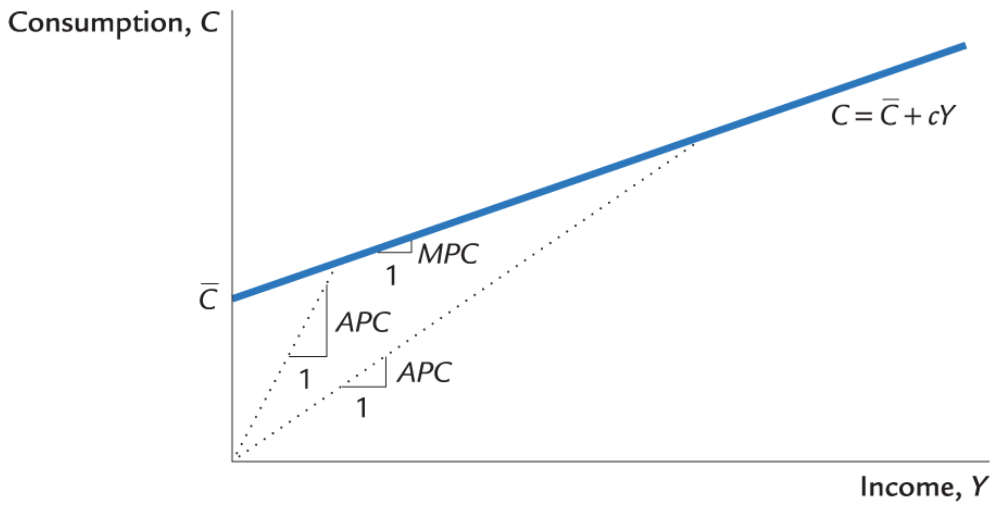
Second, Keynes posited that the ratio of consumption to income, called the **average propensity to consume**, falls as income rises. He believed that saving was a luxury, so he expected the rich to save a higher proportion of their income than the poor. Although not essential for Keynes's own analysis, the postulate that the average propensity to consume falls as income rises became a central part of early Keynesian economics.

Third, Keynes thought that income is the primary determinant of consumption and that the interest rate does not have an important role. This conjecture stood in stark contrast to the beliefs of the classical economists who preceded him. The classical economists held that a higher interest rate encourages saving and discourages consumption. Keynes admitted that the interest rate could affect consumption as a matter of theory. Yet he wrote that "the main conclusion suggested by experience is, I think, that the short-period influence of the rate of interest on individual spending out of a given income is secondary and relatively unimportant."

To express these ideas mathematically, the Keynesian consumption function is written as

C=C¯+cY, C¯>0, 0<c<1, $$C = \overline{C} + cY, \qquad \overline{C} > 0,\ 0 < c < 1,$$

where *C* is consumption, *Y* is disposable income, C¯ $\overline{C}$ is a constant, and *c* is the marginal propensity to consume. This consumption function, shown in Figure 19-1, is graphed as a straight line. C¯ $\overline{C}$ determines the intercept on the vertical axis, and *c* determines the slope.



Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

FIGURE 19-1 The Keynesian Consumption Function This figure graphs a consumption function with the three properties that Keynes conjectured. First, the marginal propensity to consume *c* is between zero and one. Second, the average propensity to consume falls as income rises. Third, consumption is determined by current income.

Note: The marginal propensity to consume *MPC* is the slope of the consumption function. The average propensity to consume APC=C/Y $APC = C/Y$ equals the slope of a line drawn from the origin to a point on the consumption function.

This consumption function exhibits the three properties that Keynes posited. It satisfies Keynes's first property because the marginal propensity to consume *c* is between zero and one, so that higher income leads to higher consumption and higher saving. This consumption function satisfies Keynes's second property because the average propensity to consume *APC* is

$$APC = C/Y = \overline{C}/Y + c.$$

As *Y* rises, $\overline{C}/Y$ falls, and so the average propensity to consume C/Y falls. Finally, this consumption function satisfies Keynes's third property because the interest rate is not included in this equation as a determinant of consumption.

The Early Empirical Successes

Soon after Keynes proposed the consumption function, economists began collecting and examining data to test his conjectures. The earliest studies indicated that the Keynesian consumption function was a good approximation of how consumers behave.

In some of these studies, researchers surveyed households and collected data on consumption and income. They found that households with higher income consumed more, which confirms that the marginal propensity to consume is greater than zero. They also found that households with higher income saved more, which confirms that the marginal propensity to consume is less than one. In addition, these researchers found that higher-income households saved a larger fraction of their income, which confirms that the average propensity to consume falls as income rises. Thus, these data verified Keynes's conjectures about the marginal and average propensities to consume.

In other studies, researchers examined aggregate data on consumption and income for the period between the two world wars. These data also supported the Keynesian consumption function. In years when income was unusually low, such as during the depths of the Great Depression in 1932 and 1933, both consumption and saving were low, indicating that the marginal propensity to consume is between zero and one. In addition, during those years of low income, the ratio of consumption to income was high, confirming Keynes's second conjecture. Finally, because the correlation between income and consumption was so strong, no other variable appeared to be important for explaining consumption. Thus, the data also confirmed Keynes's third conjecture that income is the primary determinant of how much people choose to consume and that the interest rate plays a minor role.

The Consumption Puzzle

Although the Keynesian consumption function had some early successes, two anomalies soon arose. Both concern Keynes's conjecture that the average propensity to consume falls as income rises.

The first anomaly became apparent after some economists made a dire—and, it turned out, erroneous—prediction during World War II. Based on the Keynesian consumption function, these economists reasoned that as incomes in the economy grew over time, households would consume a declining fraction of their incomes and save an increasing fraction. They feared that there might not be enough profitable investment projects to absorb all this saving. If that was indeed the case, the low consumption would lead to an inadequate demand for goods and services, resulting in a depression once the wartime demand from the government ceased. In other words, using the Keynesian consumption function, these economists predicted that the economy would experience what they called *secular stagnation*—a long depression of indefinite duration—unless the government used fiscal policy to expand aggregate demand.

Fortunately for the economy, but unfortunately for the Keynesian consumption function, the end of World War II did not throw the country into another depression. Although incomes were much higher after the war than before, these higher incomes did not lead to large increases in the rate of saving. Keynes's conjecture that the average propensity to consume would fall as income rose appeared not to hold.

The second anomaly arose when economist Simon Kuznets constructed new aggregate data on consumption and income dating back to 1869. Kuznets assembled these data in the 1940s and would later receive the Nobel Prize for this work. He discovered that the ratio of consumption to income was remarkably stable from decade to decade, despite large increases in income over the period he studied. Again, Keynes's conjecture that the average propensity to consume would fall as income rose appeared not to hold.

The failure of the secular-stagnation hypothesis and the findings of Kuznets both indicated that the average propensity to consume is fairly constant over long periods of time. This fact presented a puzzle that motivated much of the subsequent research on consumption. Economists wanted to know why some studies confirmed Keynes's conjectures and others refuted them. That is, why did Keynes's conjectures hold up well in the studies of household data and in the studies of short time-series but fail when long time-series were examined?

Figure 19-2 illustrates the puzzle. The evidence suggested that there were two consumption functions. For the household data and for the short time-series, the Keynesian consumption function appeared to work well. Yet for the long time-series, the consumption function appeared to exhibit a constant average propensity to consume. In Figure 19-2, these two relationships between consumption and income are called the short-run and long-run consumption functions. Economists needed to explain how these two consumption functions could be consistent with each other.

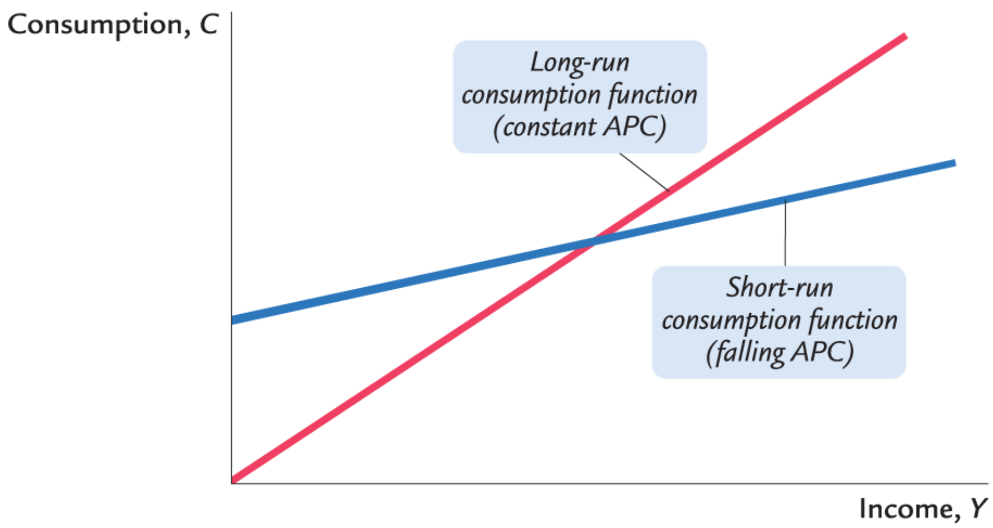


FIGURE 19-2 The Consumption Puzzle Studies of household data and short time-series found a relationship between consumption and income similar to the one Keynes conjectured. In the figure, this relationship is called the short-run consumption function. But studies of long time-series found that the average propensity to consume did not vary systematically with income. This relationship is called the long-run consumption function. Note that the short-run consumption function has a falling average propensity to consume, whereas the long-run consumption function has a constant average propensity to consume.

In the 1950s, Franco Modigliani and Milton Friedman each proposed explanations of these seemingly contradictory findings. Both economists later won Nobel Prizes, in part for their work on consumption. Modigliani and Friedman started with the same insight: *If people prefer consumption to be smooth year to year rather than widely fluctuating, they should be forward-looking. Their spending should depend not only on their current income but also on the income they expect to receive in the future*. But the two economists took this insight in different directions.

Franco Modigliani and the Life-Cycle Hypothesis

In a series of papers written in the 1950s, Franco Modigliani and his collaborators tried to solve the consumption puzzle—that is, to explain the apparently conflicting pieces of evidence that came to light when Keynes's consumption function was confronted with the data. If consumers are forward looking, Modigliani reasoned, consumption should depend on a person's lifetime income. Yet income varies systematically over people's lives. Saving allows consumers to move income from those times in life when income is high to those times when it is low. This interpretation of consumer behavior formed the basis for his **life-cycle hypothesis**.[1]

The Hypothesis

One important reason that income varies over a person's life is retirement. Most people plan to stop working at about age 65, and they expect their incomes to fall when they retire. Yet they do not want a large drop in their standard of living, as measured by their consumption. To maintain consumption after retirement, people must save during their working years. Let's see what this motive for saving implies for the consumption function.

Suppose a consumer expects to live another *T* years, has wealth of *W*, and expects to earn income *Y* per year until she retires *R* years from now. What level of consumption will the consumer choose if she wants stable consumption over the course of her life?

The consumer's lifetime resources are composed of initial wealth *W* and lifetime earnings $R \times Y$. (For simplicity, we are assuming an interest rate of zero; if the interest rate were greater than zero, we would need to take account of interest earned on savings as well.) The consumer can divide up her lifetime resources among her *T* remaining years of life. To achieve the smoothest possible path of consumption over her lifetime, she divides this total of $W + RY$ equally among the *T* years and each year consumes

$$C = (W + RY)/T.$$

We can write this person's consumption function as

$$C = (1/T)W + (R/T)Y.$$

For example, if the consumer expects to live for 50 more years and work for 30 of them, then $T = 50$ and $R = 30$, so her consumption function is

$$C = 0.02W + 0.6Y.$$

This equation says that consumption depends on both income and wealth. An extra $1 of income per year raises consumption by $0.60 per year, and an extra $1 of wealth raises consumption by $0.02 per year.

If everyone plans consumption like this, the aggregate consumption function is much the same as the individual one: aggregate consumption depends on both wealth and income. That is, the economy's consumption function is

$$C = \alpha W + \beta Y,$$

where the parameter α is the marginal propensity to consume out of wealth and the parameter β is the marginal propensity to consume out of income.

Implications

Figure 19-3 graphs the relationship between consumption and income predicted by the life-cycle model. For any given wealth *W*, the model yields a conventional consumption function similar to the one shown in Figure 19-1. Notice, however, that the intercept of the consumption function, which shows what would happen to consumption if income ever fell to zero, is not a fixed value, as it is in Figure 19-1. Instead, the intercept here is αW and, thus, depends on wealth.

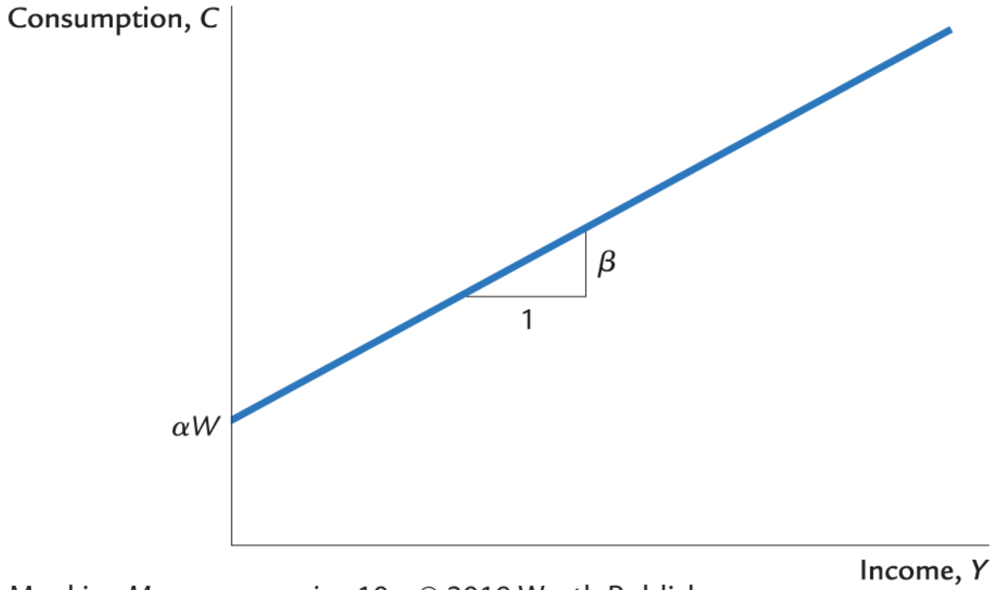


FIGURE 19-3 The Life-Cycle Consumption Function The life-cycle model says that consumption depends on wealth as well as income. As a result, the intercept of the consumption function αW depends on wealth.

This life-cycle model of consumer behavior can solve the consumption puzzle. According to the life-cycle consumption function, the average propensity to consume is

$$C/Y = \alpha(W/Y) + \beta.$$

Because wealth does not vary proportionately with income from person to person or from year to year, we should find that high income corresponds to a low average propensity to consume when we look at data across individuals or over short periods of time. But over long periods of time, wealth and income grow together, resulting in a constant ratio W/Y and thus a constant average propensity to consume.

To make the same point somewhat differently, consider how the consumption function changes over time. As Figure 19-3 shows, for any given wealth, the life-cycle consumption function looks like the one Keynes suggested. But this function holds only in the short run when wealth is constant. In the long run, as wealth increases, the consumption function shifts upward, as in Figure 19-4. This upward shift prevents the average propensity to consume from falling as income increases. In this way, Modigliani resolved the consumption puzzle posed by Simon Kuznets's data.

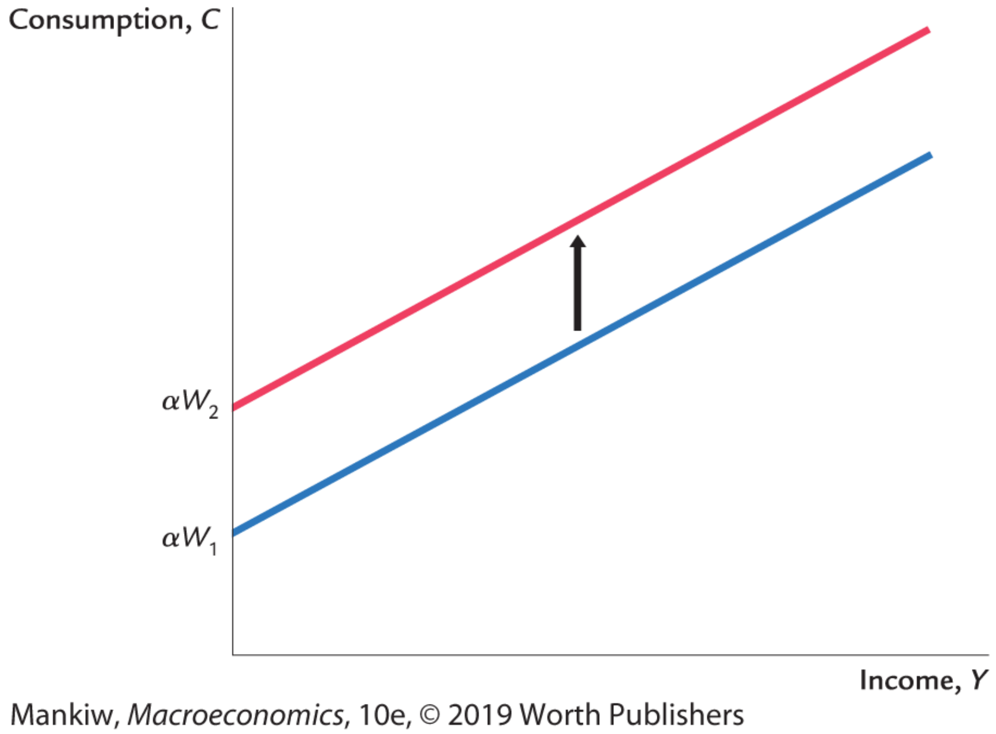


FIGURE 19-4 How Changes in Wealth Shift the Consumption Function If consumption depends on wealth, then an increase in wealth shifts the consumption function upward. Thus, the short-run consumption function (which holds wealth constant) will not continue to hold in the long run (as wealth rises over time).

The life-cycle model makes many other predictions as well. Most importantly, it predicts that saving varies over a person's lifetime. If a person begins adulthood with no wealth, she will accumulate wealth during her working years and then run down her wealth during her retirement years. Figure 19-5 shows the consumer's income, consumption, and wealth over her adult life. According to the life-cycle hypothesis, because people want to smooth consumption over their lives, the young who are working save, while the old who are retired dissave.

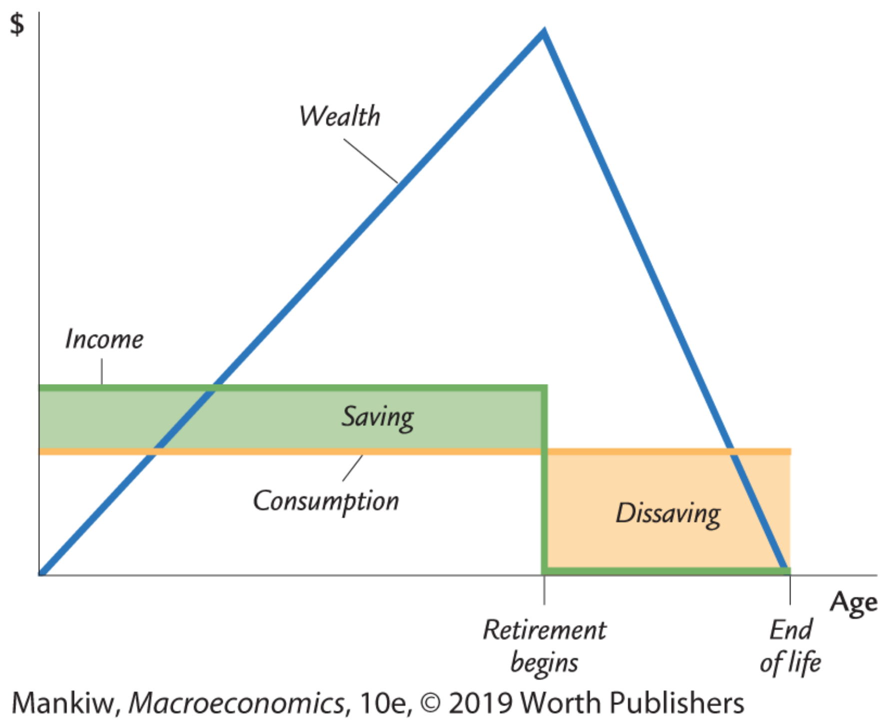


FIGURE 19-5 Consumption, Income, and Wealth over the Life Cycle If the consumer smooths consumption over her life (as indicated by the horizontal consumption line), she will save and accumulate wealth during her working years and then dissave and run down her wealth during retirement.

Motivated by this model, many economists have studied the consumption and saving of the elderly. They often find that the elderly do not dissave as much as the model predicts. In other words, the elderly do not run down their wealth as quickly as one would expect if they were smoothing their consumption over their remaining years of life. One reason may be the uncertainty the elderly face regarding life span and future medical expenses. Another reason may be that they want to leave bequests to their descendants. Providing for retirement is one motive for saving, but other motives appear to be important as well.[2]

Milton Friedman and the Permanent-Income Hypothesis

In a book published in 1957, Milton Friedman proposed the **permanent-income hypothesis** to explain consumer behavior. Friedman's permanent-income hypothesis complements Modigliani's life-cycle hypothesis: both argue that consumption should not depend on current income alone. But unlike the life-cycle hypothesis, which emphasizes that income follows a regular pattern over a person's lifetime, the permanent-income hypothesis emphasizes that people experience random changes in their incomes from year to year.[3]

The Hypothesis

Friedman suggested that we view current income Y as the sum of two components, **permanent income** YP

Y^P and **transitory income** Y^T. That is,

$$Y = Y^P + Y^T.$$

Permanent income is the part of income that people expect to persist into the future. Transitory income is the part of income that people do not expect to persist. Put differently, permanent income is average income, and transitory income is the random deviation from that average.

To see how we might separate income into these two parts, consider these examples:

- Maria, who has a law degree, earned more this year than John, who is a high school dropout. Maria's higher income resulted from higher permanent income because her education will continue to provide her a higher salary.
- Sue, a Florida orange grower, earned less than usual this year because a freeze destroyed her crop. Julio, a California orange grower, earned more than usual because the freeze in Florida drove up the price of oranges. Julio's higher income resulted from higher transitory income because he is no more likely than Sue to have good weather next year.

These examples show that different forms of income have different degrees of persistence. A good education provides a permanently higher income, whereas good weather provides only transitorily higher income. Although one can imagine intermediate cases, it is useful to keep things simple by supposing that there are only two kinds of income: permanent and transitory.

Friedman reasoned that consumption should depend primarily on permanent income because consumers use saving and borrowing to smooth consumption in response to transitory changes in income. For example, if a person received a permanent raise of $10,000 per year, her consumption would rise by about as much. Yet if a person won $10,000 in a lottery, she would not consume it all in one year. Instead, she would spread the extra consumption over the rest of her life. If we assume an interest rate of zero and a remaining life span of fifty years, consumption would rise by only $200 per year in response to the $10,000 prize. Thus, consumers spend their permanent income, but they save most of their transitory income.

Friedman concluded that we should view the consumption function as approximately

$$C = \alpha Y^P,$$

where α is a constant that measures the fraction of permanent income consumed. The permanent-income hypothesis, as expressed by this equation, states that consumption is proportional to permanent income.

Implications

The permanent-income hypothesis solves the consumption puzzle by suggesting that the standard Keynesian consumption function uses the wrong variable. According to the permanent-income hypothesis, consumption depends on permanent income Y^P; yet many studies of the consumption function try to relate consumption to current income *Y*. Friedman argued that this *errors-in-variables problem* explains the seemingly contradictory findings.

Let's see what Friedman's hypothesis implies for the average propensity to consume. Divide both sides of his consumption function by *Y* to obtain

$$APC = C/Y = \alpha Y^P/Y.$$

According to the permanent-income hypothesis, the average propensity to consume depends on the ratio of permanent income to current income. When current income temporarily rises above permanent income, the average propensity to consume temporarily falls; when current income temporarily falls below permanent income, the average propensity to consume temporarily rises.

Now consider the studies of household data. Friedman reasoned that these data reflect a combination of permanent and transitory income. Households with high permanent income have proportionately higher consumption. If all variation in household income came from the permanent component, the average propensity to consume would be the same in all households. But some of the variation in income comes from the transitory component, and households with high transitory income do not have higher consumption. Therefore, researchers find that high-income households have, on average, lower average propensities to consume.

Similarly, consider the studies of time-series data. Friedman reasoned that year-to-year fluctuations in income are dominated by transitory income. Therefore, years of high income should be years of low average propensities to consume. But over long periods of time—say, from decade to decade—the variation in income comes from the permanent component. Hence, in long time-series, one should observe a constant average propensity to consume, as in fact Kuznets found.

CASE STUDY

The 1964 Tax Cut and the 1968 Tax Surcharge

The permanent-income hypothesis can help us interpret how the economy responds to changes in fiscal policy. According to the *IS–LM* model described in Chapters 11 and 12, tax cuts stimulate consumption and raise aggregate demand, and tax increases depress consumption and reduce aggregate demand. The permanent-income hypothesis, however, predicts that consumption responds only to changes in permanent income. Therefore, transitory changes in taxes should have only a negligible effect on consumption and aggregate demand.

That's the theory. But one might naturally ask: is this prediction actually borne out in the data?

Some economists say yes, and they point to two historical changes in fiscal policy—the tax cut of 1964 and the tax surcharge of 1968—to illustrate the principle. The tax cut of 1964 was popular. It was announced as being a major and permanent reduction in tax rates. As we discussed in Chapter 11, this policy change had the intended effect of stimulating the economy.

The tax surcharge of 1968 arose in a very different political climate. It became law because the economic advisers of President Lyndon Johnson believed that the increase in government spending from the Vietnam War had excessively stimulated aggregate demand. To offset this effect, they recommended a tax increase. But Johnson, aware that the war was already unpopular, feared the political repercussions of higher taxes. He agreed to a temporary tax surcharge—in essence, a one-year increase in taxes. The tax surcharge did not seem to have the desired effect of reducing aggregate demand. Unemployment continued to fall, and inflation continued to rise. This is what the permanent-income hypothesis predicts: the tax increase affected only transitory income, so consumption behavior and aggregate demand were not greatly affected.

While these two historical examples are consistent with the permanent-income hypothesis, we cannot draw firm inferences from them. At any moment in time, there are many influences on consumer spending, including the overall confidence that consumers have in their prospects. It is hard to disentangle the effects of tax policy from the effects of other events occurring at the same time. Fortunately, some recent research has reached more reliable conclusions, as we discuss next. ■

CASE STUDY

The Tax Rebates of 2008

When medical researchers want to know the effectiveness of a new treatment, the best approach is to conduct a randomized controlled experiment. A group of patients is assembled. Half of them are given the new treatment, and the rest are given a placebo. The researchers can then track and compare the two groups to measure the effects of the treatment.

Macroeconomists usually cannot conduct randomized experiments, but sometimes such experiments fall into our lap as accidents of history. One example occurred in 2008. Because of the financial crisis that year, the economy was heading into a recession. To counteract the recessionary forces, Congress passed the Economic Stimulus Act, which provided $100 billion of one-time tax rebates to households. Single individuals received $300 to $600, couples received $600 to $1,200, and families with children received an additional $300 per child. Most importantly, because sending out many millions of checks was a lengthy process, consumers received their tax rebates at different times. The timing of receipt was based on the last two digits of the individual's Social Security number, which is essentially random. By comparing the spending behavior of consumers who received early payments to the behavior of those who received later payments, researchers could use this random variation to estimate the effect of a transitory tax cut.

Here are the results, as reported by the researchers who did the study:

> We find that on average households spent about 12 to 30 percent of their stimulus payments, depending on the specification, on nondurable consumption goods and services (as defined in the consumer expenditure survey) during the three-month period in which the payments were received. This response is statistically and economically significant. We also find a significant effect on the purchase of durable goods and related services, primarily the purchase of vehicles, bringing the average response of total consumption expenditures to about 50 to 90 percent of the payments during the three-month period of receipt.[4]

The findings of this study stand in stark contrast to what the permanent-income hypothesis predicts. If households were smoothing their consumption over time, as the permanent-income hypothesis assumes, they would have spent only a small fraction of the tax rebate in a three-month period, but the data show a large impact of the rebate on spending. In addition, if the permanent-income hypothesis were correct, those receiving the early checks would not have behaved any differently from those receiving the later checks because the permanent income of the two groups was the same. Yet the data show that the timing of the check's arrival had a profound impact on the timing of a household's spending.

One possible explanation for these findings is that many households face **borrowing constraints**—limits on the amount they can borrow against expected future income. Friedman's permanent income hypothesis is based on the premise that households can use saving and borrowing to smooth consumption over time. Borrowing constraints impede consumption smoothing and tie a household's spending to its current income.

The permanent-income theory may be right that permanent tax changes influence consumer spending more powerfully than transitory ones. But based on the evidence from the 2008 experience, it seems wrong to conclude that the effects of transitory tax changes are insignificantly small. Even very transitory changes in tax policy can influence how much consumers spend. ■

Robert Hall and the Random-Walk Hypothesis

The permanent-income hypothesis is built on the insight that forward-looking consumers base their consumption decisions not only on their current income but also on their expected future income. Thus, the permanent-income hypothesis highlights the idea that consumption depends on people's expectations.

Subsequent research on consumption combined this view of the consumer with the assumption of rational expectations. The rational-expectations assumption states that people use all available information to make optimal forecasts about the future. As we saw in Chapter 14, this assumption can have profound implications for the costs of stopping inflation. It can also have profound implications for the study of consumer behavior.

The Hypothesis

The economist Robert Hall was the first to derive the implications of rational expectations for consumption.

He showed that if the permanent-income hypothesis is correct, and if consumers have rational expectations, then changes in consumption over time should be unpredictable. When changes in a variable are unpredictable, the variable is said to follow a **random walk**. According to Hall, the combination of the permanent-income hypothesis and rational expectations implies that consumption follows a random walk.

Hall reasoned as follows. According to the permanent-income hypothesis, consumers face fluctuating income and try their best to smooth their consumption over time. At any moment, consumers choose consumption based on their current expectations of their lifetime incomes. Over time, they change their consumption because they receive news that causes them to revise their expectations. For example, a person getting an unexpected promotion increases consumption, whereas a person getting an unexpected demotion decreases consumption. In other words, changes in consumption reflect “surprises” about lifetime income. If consumers are optimally using all available information, they should be surprised only by events that were unpredictable. Therefore, changes in their consumption should be unpredictable as well.[5]

Implications

The rational-expectations approach to consumption has implications not only for forecasting but also for the analysis of economic policies. *If consumers obey the permanent-income hypothesis and have rational expectations, only unexpected policy changes influence consumption. These policy changes take effect when they change expectations.* For example, suppose Congress passes a tax increase to be effective next year. In this case, consumers receive the news about their lifetime incomes when Congress passes the law (or even earlier, if the law’s passage was predictable). The arrival of this news causes consumers to revise their expectations and reduce their consumption. The following year, when the tax hike goes into effect, consumption is unchanged because no news has arrived.

Hence, if consumers have rational expectations, policymakers influence the economy not only through their actions but also through the public’s expectation of their actions. Expectations, however, cannot be observed directly. Therefore, it is often hard to know how and when changes in fiscal policy alter aggregate demand.

CASE STUDY

Do Predictable Changes in Income Lead to Predictable Changes in Consumption?

Of the many facts about consumer behavior, one is impossible to dispute: income and consumption fluctuate together over the business cycle. When the economy goes into a recession, both income and consumption fall, and when the economy booms, both income and consumption rise rapidly.

By itself, this fact doesn’t say much about the rational-expectations version of the permanent-income hypothesis. Most short-run fluctuations are unpredictable. Thus, when the economy goes into a recession, the typical consumer is receiving bad news about her lifetime income, so consumption naturally falls. And when the economy booms, the typical consumer is receiving good news about her lifetime income, so consumption rises. This behavior does not necessarily violate the random-walk theory that changes in consumption are impossible to forecast.

Yet suppose we could identify some *predictable* changes in income. According to the random-walk theory, these changes in income should not cause consumers to revise their spending plans. If consumers expected income to rise or fall, they should have adjusted their consumption already in response to that information. Thus, predictable changes in income should not lead to predictable changes in consumption.

Data on consumption and income, however, do not satisfy this implication of the random-walk theory. When income is expected to fall by $1, consumption will on average fall at the same time by about $0.50. In other words, predictable changes in income lead to predictable changes in consumption that are roughly half as large.

Why is this so? One possible explanation of this behavior is that some consumers may fail to have rational expectations. Instead, they may base their expectations of future income excessively on current income. Thus, when income rises or falls (even predictably), they act as if they received news about their lifetime resources and change their consumption accordingly. Another possible explanation is that some consumers are borrowing-constrained and, therefore, base their consumption on current income alone. Regardless of which explanation is correct, Keynes's original consumption function starts to look more attractive. That is, current income has a larger role in determining consumer spending than the random-walk hypothesis suggests.[6] ■

David Laibson and the Pull of Instant Gratification

Keynes called the consumption function a "fundamental psychological law." Yet psychology did not play a large role in the subsequent study of consumption. Most economists assumed that consumers are rational maximizers of utility who are always evaluating their opportunities and plans to obtain the highest lifetime satisfaction. Modigliani, Friedman, and Hall all relied on this model of human behavior as they developed their theories of consumption.

More recently, economists have returned to psychology. They have suggested that consumption decisions are not made by the ultrarational *Homo economicus* but by real human beings whose behavior is more complex. This new subfield infusing psychology into economics is called *behavioral economics*. The most prominent behavioral economist studying consumption is David Laibson.

Laibson notes that many consumers judge themselves to be imperfect decisionmakers. In one survey of the American public, 76 percent said they were not saving enough for retirement. In another survey of the baby-boom generation, respondents were asked the percentage of income that they save and the percentage that they thought they should save. The saving shortfall averaged 11 percentage points.

According to Laibson, the insufficiency of saving is related to another phenomenon: the pull of instant gratification. Consider the following two questions:

Question 1: Would you prefer (A) a candy today or (B) two candies tomorrow?

Question 2: Would you prefer (A) a candy in 100 days or (B) two candies in 101 days?

Many people will answer A to the first question and B to the second. In a sense, they are more patient in the long run than they are in the short run.

This raises the possibility that consumers may have **time-inconsistent preferences**: they may alter their decisions simply because time passes. A person confronting question 2 may choose B and wait the extra day for the extra candy. But after 100 days pass, she finds herself in a new short run, confronting question 1. The pull of instant gratification may induce her to change her mind.

We see this kind of behavior in many situations in life. A person on a diet may have a second helping at dinner, while promising herself that she will eat less tomorrow. A person may smoke one more cigarette, while promising herself that this is the last one. And a consumer may splurge at the shopping mall, while promising herself that tomorrow she will cut back her spending and start saving more for retirement. But when tomorrow arrives, the promises are in the past, and a new self takes control of the decisionmaking, with its own desire for instant gratification.

The possibility that consumers may deviate from conventional rationality and exhibit time-inconsistent behavior is potentially important for designing public policies, as the following Case Study discusses.[7]

CASE STUDY

How to Get People to Save More

Many economists believe that it would be desirable for Americans to increase the fraction of their income that they save. There are several reasons for this conclusion. From a microeconomic perspective, greater saving would mean that people would be better prepared for retirement; this goal is especially important because Social Security, the public program that provides retirement income, is projected to run into financial difficulties in the years ahead as the population ages. From a macroeconomic perspective, greater saving would increase the supply of loanable funds available to finance investment; the Solow growth model shows that increased capital accumulation leads to higher income. From an open-economy perspective, greater saving would mean that less domestic investment would be financed by capital flows from abroad; a smaller capital inflow pushes the trade balance from deficit toward surplus. Finally, the fact that many Americans say that they are not saving enough may be sufficient reason to think that increased saving should be a national goal.

The difficult issue is how to get Americans to save more. The burgeoning field of behavioral economics offers some answers.

One approach is to make saving the path of least resistance. For example, consider 401(k) plans, the tax-advantaged retirement savings accounts available to many workers through their employers. In most firms, participation in the plan is an option that workers can choose by filling out a simple form. In some firms, however, workers are automatically enrolled in the plan but can opt out by filling out a simple form. Studies have shown that workers are far more likely to participate in the second case than in the first. If workers were rational maximizers, as is often assumed in economic theory, they would choose the optimal amount of saving, regardless of whether they had to choose to enroll or were enrolled automatically. In fact, because workers exhibit inertia, the default has a powerful influence over how much they save. Policymakers who want to increase saving can take advantage of this inertia by making automatic enrollment more common.

A second approach to increasing saving is to give people the opportunity to control their desires for instant gratification. That is the goal of the “Save More Tomorrow” program proposed by economist Richard Thaler, who won the Nobel Prize in 2017. The essence of this program is that people commit in advance to putting a portion of their future salary increases into a retirement savings account. When a worker signs up, she makes no sacrifice of lower consumption today but, instead, commits to reducing consumption growth in the future. When this plan was implemented in several firms, it had a large impact. A high proportion (78 percent) of those offered the plan joined. In addition, of those enrolled, the vast majority (80 percent) stayed with the program through at least the fourth annual pay raise. The average saving rates for those in the program increased from 3.5 percent to 13.6 percent over the course of forty months.

How successful would wider application of these ideas be in increasing the U.S. national saving rate? It is hard to know. But given the importance of saving to both personal and national prosperity, many economists believe these proposals are worth a try.[8] ■

The Bottom Line on Consumption

In the work of five economists, we have seen a range of views on consumer behavior. Keynes proposed that consumption depends largely on current income. He suggested a consumption function of the form

Consumption=f(Current Income).

$$\text{Consumption} = f(\text{Current Income}).$$

More recently, economists have argued that consumers look ahead to their future resources and needs, implying a more complex consumption function than the one Keynes proposed. This work suggests instead that

Consumption = f(Current Income, Wealth, Expected Future Income, Interest Rates, Self-Control Mechanisms).

$$\text{Consumption} = f(\text{Current Income, Wealth, Expected Future Income,} \\ \text{Interest Rates, Self-Control Mechanisms}).$$

In other words, current income is only one determinant of aggregate consumption.

Economists debate the importance of these determinants of consumption. There remains disagreement about, for example, the influence of interest rates on consumer spending, the prevalence of borrowing constraints, and the importance of psychological effects. Economists sometimes disagree about economic policy because they assume different consumption functions.

19-2 What Determines Investment Spending?

While spending on consumption goods provides utility to households today, spending on investment goods is aimed at providing a higher standard of living at a later date. Investment is the component of GDP that links the present and the future.

Investment spending is also the most volatile component of GDP. When expenditure on goods and services falls during a recession, much of the decline is usually due to a drop in investment. In the Great Recession of 2008–2009, for example, U.S. real GDP fell $636 billion from its peak in the fourth quarter of 2007 to its trough in the second quarter of 2009. Investment spending over the same period fell $785 billion, accounting for more than the entire fall in spending.

As we saw in Chapter 2, there are three types of investment spending: business fixed investment, residential investment, and inventory investment. Here we focus on business fixed investment, which accounts for about three-quarters of investment spending. The term *business* means that these capital goods are bought by firms for use in future production. The term *fixed* means that this spending is for capital that will stay put for a while, as opposed to inventory investment, which will be used or sold within a short time. Business fixed investment includes everything from office furniture to factories, computers to company cars.

The standard model of business fixed investment is called the **neoclassical model of investment**. The neoclassical model examines the benefits and costs to firms of owning capital goods. The model shows how investment—the addition to the stock of capital—is related to the marginal product of capital, the interest rate, and the tax rules affecting firms.

To develop the model, we will imagine that there are two kinds of firms. *Production firms* produce goods and services using capital that they rent (just as in our model in Chapter 3). *Rental firms* make all the investments in the economy; they buy capital and rent it out to the production firms. In actuality, most real firms both produce goods and services and invest in capital for future production. We can clarify our thinking, however, if we separate these two activities by imagining that they take place in different firms.

The Rental Price of Capital

Let's first consider the typical production firm. As we saw in Chapter 3, this firm decides how much capital to rent by comparing the cost and benefit of each unit of capital. The firm rents capital at a rental rate *R* and sells its output at a price P; the real cost of a unit of capital to the production firm is R/P. The real benefit of a unit of capital is the marginal product of capital *MPK*—the extra output produced with one more unit of capital. The marginal product of capital declines as the amount of capital rises: the more capital the firm has, the less an additional unit of capital will add to its output. Chapter 3 concluded that, to maximize profit, the firm rents capital until the marginal product of capital falls to equal the real rental price.

Figure 19-6 shows the equilibrium in the rental market for capital. For the reasons just discussed, the marginal product of capital determines the demand curve. The demand curve slopes downward because the marginal product of capital is low when the level of capital is high. At any point in time, the amount of capital in the economy is fixed, so the supply curve is vertical. The real rental price of capital adjusts to equilibrate supply and demand.

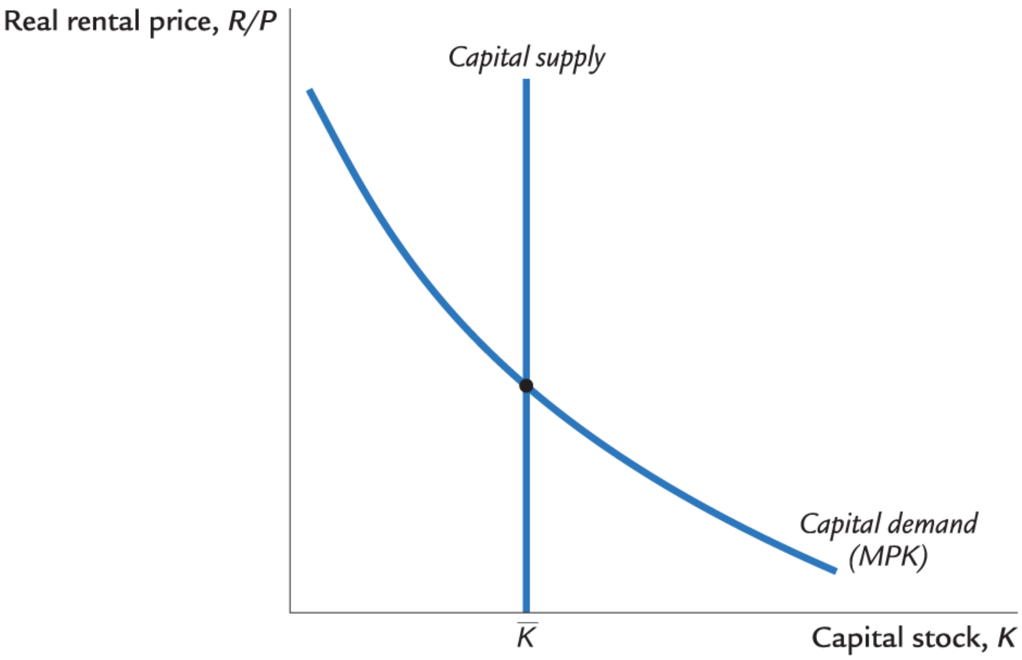


FIGURE 19-6 The Rental Price of Capital The real rental price of capital adjusts to equilibrate the demand for capital (determined by the marginal product of capital) and the fixed supply.

To see what variables influence the equilibrium rental price, let's consider a particular production function. As we saw in Chapter 3, many economists consider the Cobb–Douglas production function a good approximation of how the economy turns capital and labor into goods and services. The Cobb–Douglas production function is

$$Y = AK^{\alpha}L^{1-\alpha},$$

where Y is output, K is capital, L is labor, A is a parameter measuring the level of technology, and α is a parameter between zero and one that measures capital's share of output. The marginal product of capital for the Cobb–Douglas production function is

$$MPK = \alpha A(L/K)^{1-\alpha}.$$

Because the real rental price R/P equals the marginal product of capital in equilibrium, we can write

$$R/P = \alpha A(L/K)^{1-\alpha}.$$

This expression identifies the variables that determine the real rental price. It shows the following:

- The lower the capital stock, the higher the real rental price of capital.
- The greater the amount of labor employed, the higher the real rental price of capital.
- The better the technology, the higher the real rental price of capital.

Events that reduce the capital stock (a tornado), raise employment (an expansion in aggregate demand), or improve the technology (a scientific discovery) raise the equilibrium real rental price of capital.

The Cost of Capital

Next consider the rental firms. These firms, like car-rental companies, buy capital goods and rent them out. Because our goal is to explain the investments made by the rental firms, we begin by considering the benefit and cost of owning capital.

The benefit of owning capital is the revenue earned by renting it to the production firms. The rental firm receives the real rental price of capital R/P for each unit of capital it owns and rents out.

The cost of owning capital is more complex. For each period of time that it rents out a unit of capital, the rental firm bears three costs:

1. When a rental firm borrows to buy a unit of capital, it must pay interest on the loan. If P_K is the purchase price of a unit of capital and *i* is the nominal interest rate, then iP_K is the interest cost. Notice that this interest cost would be the same even if the rental firm did not have to borrow: if the rental firm buys a unit of capital using cash on hand, it loses out on the interest it could have earned by depositing this cash in the bank. In either case, the interest cost equals iP_K.
2. While the rental firm is renting out the capital, the price of capital can change. If the price of capital falls, the firm loses because the firm's asset has fallen in value. If the price of capital rises, the firm gains because the firm's asset has risen in value. The cost of this loss or gain is $-\Delta P_K$. (The minus sign is here because we are measuring costs, not benefits.)

3. While the capital is rented out, it suffers wear and tear, called **depreciation**. If δ is the rate of depreciation—the fraction of capital's value lost per period because of wear and tear—then the dollar cost of depreciation is δP_K.

The total cost of renting out a unit of capital for one period is therefore

$$\begin{aligned}\text{Cost of Capital} &= iP_K - \Delta P_K + \delta P_K \\ &= P_K\left(i - \Delta P_K / P_K + \delta\right).\end{aligned}$$

The cost of capital depends on the price of capital, the interest rate, the rate at which capital prices are changing, and the depreciation rate.

For example, consider the cost of capital to a car-rental company. The company buys cars for $30,000 each and rents them out to other businesses. The company faces an interest rate *i* of 10 percent per year, so the interest cost iP_K is $3,000 per year for each car the company owns. Car prices are rising at 6 percent per year, so, excluding wear and tear, the firm gets a capital gain ΔP_K of $1,800 per year. Cars depreciate at 20 percent per year, so the loss due to wear and tear δP_K is $6,000 per year. Therefore, the company's cost of capital is

$$\begin{aligned}\text{Cost of Capital} &= \$3{,}000 - \$1{,}800 + \$6{,}000 \\ &= \$7{,}200.\end{aligned}$$

The cost to the car-rental company of keeping a car in its capital stock is $7,200 per year.

To make the expression for the cost of capital simpler and easier to interpret, we assume that the price of capital goods rises with the prices of other goods. In this case, $\Delta P_K / P_K$ equals the overall rate of inflation π. Because $i - \pi$ equals the real interest rate *r*, we can write the cost of capital as

$$\text{Cost of Capital} = P_K\left(r + \delta\right).$$

This equation states that the cost of capital depends on the price of capital, the real interest rate, and the depreciation rate.

Finally, we want to express the cost of capital relative to other goods in the economy. The **real cost of capital**—the cost of buying and renting out a unit of capital measured in units of the economy's output—is

Real Cost of Capital=(PK/P)(r+δ).

$$\text{Real Cost of Capital} = (P_K/P)(r+\delta).$$

This equation states that the real cost of capital depends on the relative price of a capital good P_K/P, the real interest rate r, and the depreciation rate δ.

The Cost-Benefit Calculus of Investment

Now consider a rental firm's decision about whether to increase or decrease its capital stock. For each unit of capital, the firm earns real revenue R/P and bears the real cost $(P_K/P)(r+\delta)$. The real profit per unit of capital is

Profit Rate=Revenue − Cost=R/P−(PK/P)(r+δ).

$$\begin{aligned}\text{Profit Rate} &= \text{Revenue} - \text{Cost} \\ &= R/P - (P_K/P)(r+\delta).\end{aligned}$$

Because the real rental price in equilibrium equals the marginal product of capital, we can write the profit rate as

Profit Rate=MPK−(PK/P)(r+δ).

$$\text{Profit Rate} = MPK - (P_K/P)(r+\delta).$$

The rental firm makes a profit if the marginal product of capital is greater than the cost of capital. It incurs a loss if the marginal product is less than the cost of capital.

We can now see the incentives that lie behind the rental firm's investment decision. The firm's decision regarding its capital stock—that is, whether to add to it or to let it depreciate—depends on whether owning and renting out capital is profitable. The change in the capital stock, called **net investment**, depends on the difference between the marginal product of capital and the cost of capital. *If the marginal product of capital exceeds the cost of capital, firms find it profitable to add to their capital stock. If the marginal product of capital falls short of the cost of capital, they let their capital stock shrink*.

We can also now see that the separation of economic activity between production and rental firms, although useful for clarifying our thinking, is not necessary for our conclusion regarding how firms choose how much to invest. For a firm that both uses and owns capital, the benefit of an extra unit of capital is the marginal product of capital, and the cost is the cost of capital. Like a firm that owns and rents out capital, this firm adds to its capital stock if the marginal product exceeds the cost of capital. Thus, we can write

ΔK=In [MPK−(PK/P)(r+δ)], $$\Delta K = I_n \ [MPK - (P_K/P)(r+\delta)],$$

where In I_n( ) is the function showing how net investment responds to the incentive to invest. How much the capital stock responds (and thus the precise form of this function) depends on how costly the adjustment process is.

We can now derive the investment function. Total spending on investment is the sum of net investment and the replacement of depreciated capital. The investment function is

I=In [MPK−(PK/P)(r+δ)]+δK. $$I = I_n \ [MPK - (P_K/P)(r+\delta)] + \delta K.$$

Investment depends on the marginal product of capital, the cost of capital, and the amount of depreciation.

This model shows why investment depends on the interest rate. A decrease in the real interest rate lowers the cost of capital. It therefore raises the amount of profit from owning capital and increases the incentive to accumulate more capital. Similarly, an increase in the real interest rate raises the cost of capital and leads firms to reduce their investment. For this reason, the investment schedule relating investment to the interest rate slopes downward, as in panel (a) of Figure 19-7.

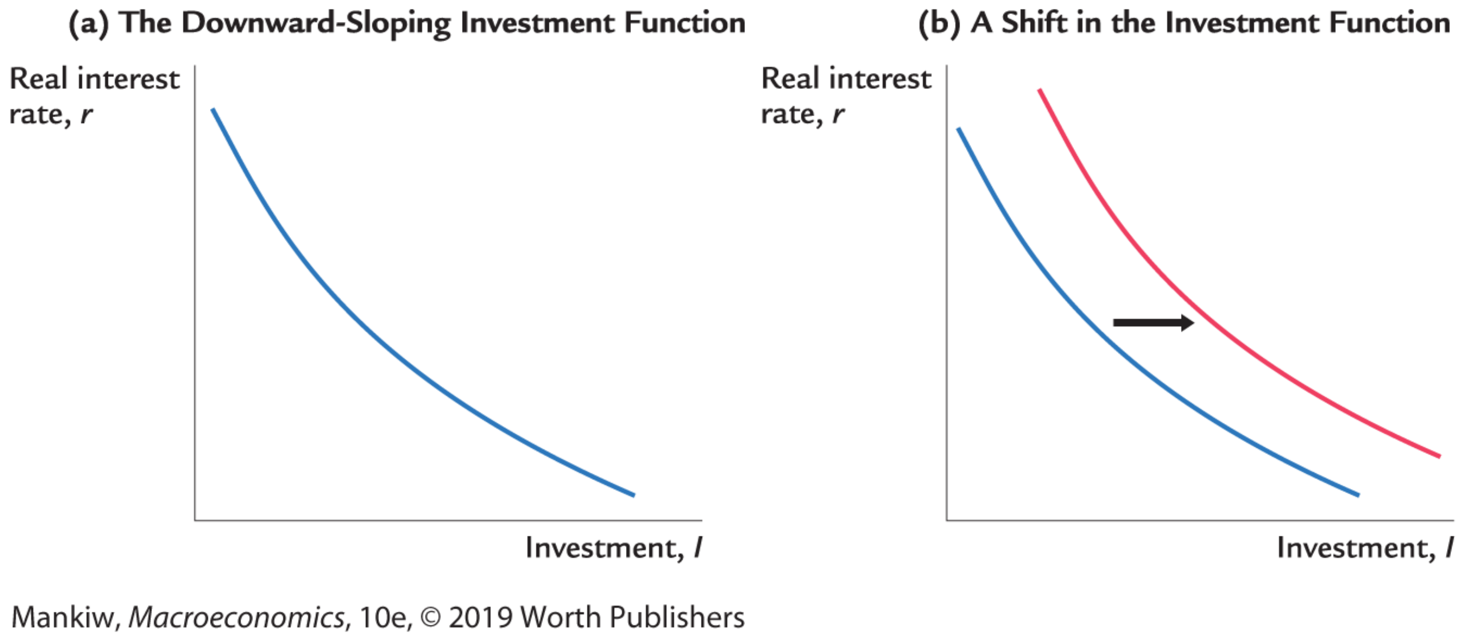


FIGURE 19-7 The Investment Function Panel (a) shows that investment increases when the interest rate falls. This is because a lower interest rate reduces the cost of capital and therefore makes owning capital more profitable. Panel (b) shows an outward shift in the investment function, which might be due to an increase in the marginal product of capital.

The model also shows what causes the investment schedule to shift. Any event that raises the marginal product of capital increases the profitability of investment and causes the investment schedule to shift outward, as in panel (b) of Figure 19-7. For example, a technological innovation that increases the production function parameter *A* raises the marginal product of capital and, for any given interest rate, increases the amount of capital goods that rental firms wish to buy.

Finally, consider what happens as this adjustment of the capital stock continues over time. If the marginal product begins above the cost of capital, the capital stock will rise and the marginal product will fall. If the marginal product of capital begins below the cost of capital, the capital stock will fall and the marginal product will rise. Eventually, as the capital stock adjusts, the marginal product of capital approaches the cost of capital. When the capital stock reaches a steady-state level, we can write

$$MPK = (P_K/P)(r+\delta).$$

Thus, in the long run, the marginal product of capital equals the real cost of capital. The speed of adjustment toward the steady state depends on how quickly firms adjust their capital stock, which in turn depends on how costly it is to build, deliver, and install new capital.[9]

Taxes and Investment

Tax laws influence firms’ incentives to accumulate capital in many ways. Sometimes policymakers change the tax code to shift the investment function and influence aggregate demand. Here we consider two of the most important provisions of corporate taxation: the corporate income tax and the investment tax credit.

The **corporate income tax** is a tax on corporate profits. Throughout much of its history, the corporate tax rate levied by the U.S. federal government was 46 percent. The rate was lowered to 34 percent in 1986 and then raised to 35 percent in 1993, and it remained at that level through 2017. Many states impose an additional corporate tax as well, bringing the total corporate tax rate in the United States to about 40 percent. By contrast, the average corporate tax rate in 2017 was 20 percent in Europe and 21 percent in Asia.At the end of 2017, to bring the United States closer to international norms, the U.S. Congress passed legislation to reduce the corporate tax rate from 35 to 21 percent, starting in 2018.

The effect of a corporate income tax on investment depends on how the tax law defines “profit.” Suppose, first, the law defined profit as we did previously—the rental price of capital minus the cost of capital. In this case, even though firms would be sharing a fraction of their profits with the government, it would still be rational for them to invest if the rental price of capital exceeded the cost of capital and to disinvest if the rental price fell short of the cost of capital. A tax on profit, measured in this way, would not alter investment incentives.

Yet, because of the tax law’s definition of profit, the corporate income tax does affect investment decisions. There are many differences between the law’s definition of profit and ours. For example, one difference is the treatment of depreciation. Our definition of profit deducts the *current* value of depreciation as a cost. That is, it bases depreciation on how much it would cost today to replace worn-out capital. By contrast, under the corporate tax laws, firms deduct depreciation using *historical* cost. That is, the depreciation deduction is based on the price of the capital when it was originally purchased. In periods of inflation, replacement cost is greater than historical cost, so the corporate tax tends to understate the cost of depreciation and overstate profit. As a result, the tax law sees a profit and levies a tax even when economic profit is zero, which makes owning capital less attractive. For this and other reasons, many economists believe that the corporate income tax discourages investment.

Policymakers often change the rules governing the corporate income tax to encourage investment or at least mitigate the disincentive the tax provides. One example is the **investment tax credit**, a tax provision that reduces a firm’s taxes by a certain amount for each dollar spent on capital goods. Because a firm recoups part of its expenditure on new capital through lower taxes, the credit reduces the effective purchase price of a unit of capital PK.P_K. Thus, the investment tax credit reduces the cost of capital and raises investment.

In 1985 the investment tax credit was 10 percent. Yet the Tax Reform Act of 1986, which reduced the corporate income tax rate, also eliminated the investment tax credit. When Bill Clinton ran for president in 1992, he campaigned on a platform of reinstituting the investment tax credit, but he did not get this proposal through Congress. The idea of reinstating the investment tax credit, however, still arises from time to time.

The tax rules regarding depreciation are another example of how policymakers can influence the incentives for investment. When George W. Bush became president, the economy was sliding into recession, attributable in large measure to a significant decline in business investment. The tax cuts Bush signed into law during his first term included provisions for temporary “bonus depreciation.” This meant that for purposes of calculating their corporate tax liability, firms could deduct the cost of depreciation earlier in the life of an investment project. This bonus, however, was available only for investments made before the end of 2004. The goal of the policy was to encourage investment at a time when the economy needed a boost to aggregate demand. According to a study by economists Christopher House and Matthew Shapiro, the goal was achieved to some degree. They write, “While their aggregate effects were probably modest, the 2002 and 2003 bonus depreciation policies had noticeable effects on the economy. For the U.S. economy as a whole, these policies may have increased GDP by $10 to $20 billion and may have been responsible for the creation of 100,000 to 200,000 jobs.” In 2011, as the economy was in the midst of the next recession, President Obama signed into law a similar measure for temporary bonus depreciation.[10]

The Stock Market and Tobin’s *q*

Many economists see a link between fluctuations in investment and fluctuations in the stock market. Recall that the term *stock* refers to shares in the ownership of corporations, and the *stock market* is the market in which these shares are traded. Stock prices tend to be high when firms have many opportunities for profitable investment because these profit opportunities mean higher future income for the shareholders. Thus, stock prices reflect the incentives to invest.

The Nobel Prize–winning economist James Tobin proposed that firms base their investment decisions on the following ratio, now called **Tobin’s *q***:

q=Market Value of Installed Capital Replacement Cost of Installed Capital .

$$q = \frac{\text{Market Value of Installed Capital}}{\text{Replacement Cost of Installed Capital}}.$$

The numerator of Tobin’s *q* is the value of the economy’s capital, as determined by the stock market. The denominator is the price of that capital if it were purchased today.

Tobin reasoned that net investment should depend on whether *q* is greater or less than 1. If *q* is greater than 1, the stock market values installed capital at more than its replacement cost. In this case, managers can raise the market value of their firms’ stock by buying more capital. Conversely, if *q* is less than 1, the stock market values capital at less than its replacement cost. In this case, managers will not replace capital as it wears out.

At first the *q* theory of investment may appear very different from the neoclassical model developed previously, but the two theories are closely related. To see the relationship, note that Tobin’s *q* depends on current and future expected profits from installed capital. If the marginal product of capital exceeds the cost of capital, then firms are earning profits on their installed capital. These profits make the firms more desirable to own, which raises the market value of these firms’ stock, implying a high value of *q*. Similarly, if the marginal product of capital falls short of the cost of capital, then firms are incurring losses on their installed capital, implying a low market value and a low value of *q*.

The advantage of Tobin’s *q* as a measure of the incentive to invest is that it reflects the expected future profitability of capital as well as the current profitability. For example, suppose Congress enacts a reduction in the corporate income tax beginning next year. This expected fall in the corporate tax means greater profits for the owners of capital. These higher expected profits raise the value of stock today, raise Tobin’s *q*, and therefore encourage investment today. Thus, Tobin’s *q* theory of investment emphasizes that investment decisions depend not only on current economic policies but also on expected future policies.[11]

CASE STUDY

The Stock Market as an Economic Indicator

"The stock market has predicted nine out of the last five recessions." So goes Paul Samuelson's famous quip about the stock market's reliability as an economic indicator. The stock market is in fact quite volatile, and it can give false signals about the future of the economy. Yet one should not ignore the link between the stock market and the economy. Figure 19-8 shows that changes in the stock market often reflect changes in real GDP. Whenever the stock market experiences a substantial decline, there is reason to fear that a recession may be around the corner.

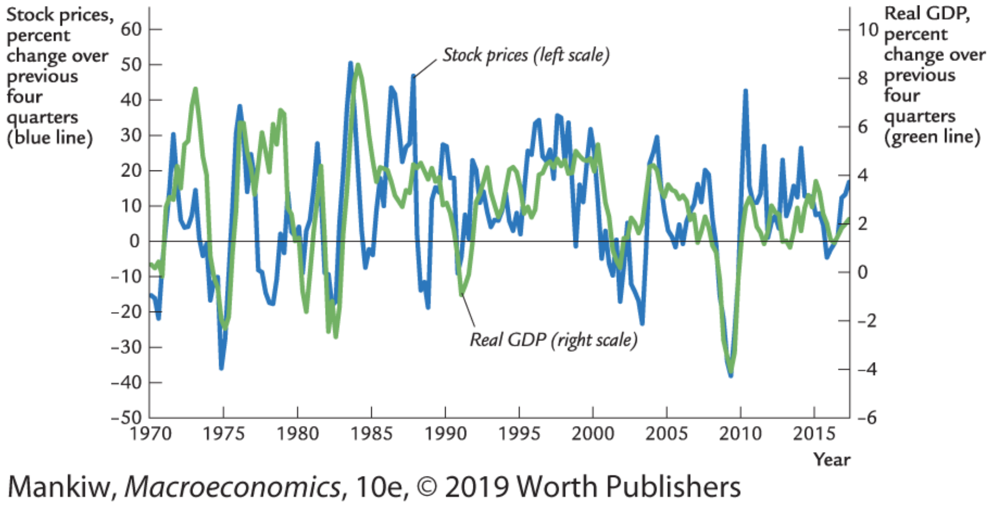


FIGURE 19-8 The Stock Market and the Economy This figure shows the association between the stock market and real economic activity. Using quarterly data from 1970 to 2017, it presents the percentage change from one year earlier in the Dow Jones Industrial Average (an index of stock prices of major industrial companies) and in real GDP. The figure shows that the stock market and GDP tend to move together but that the association is far from precise.

Data from: U.S. Department of Commerce and S&P Dow Jones Indices.

Why do stock prices and economic activity tend to fluctuate together? One reason is given by Tobin's q theory, together with the model of aggregate demand and aggregate supply. Suppose, for instance, you observe a fall in stock prices. Because the replacement cost of capital is fairly stable, a fall in the stock market is usually associated with a fall in Tobin's q. A fall in q reflects investors' pessimism about the current or future profitability of capital. This means that the investment function has shifted inward: investment is lower at any given interest rate. As a result, the aggregate demand for goods and services contracts, leading to lower output and employment.

There are two additional reasons that stock prices are associated with economic activity. First, because stock is part of household wealth, a fall in stock prices makes people poorer and thus depresses consumer spending, which also reduces aggregate demand. Second, a fall in stock prices might reflect bad news about technological progress and long-run economic growth. If so, this means that the natural level of output—and thus aggregate supply—will be growing more slowly in the future than was previously expected.

These links between the stock market and the economy are not lost on policymakers, such as those at the Federal Reserve. Indeed, because the stock market often anticipates changes in real GDP, and because data on the stock market are available more quickly than data on GDP, the stock market is a closely watched economic indicator. ■

Financing Constraints

When a firm wants to invest in new capital—say, by building a new factory—it often raises the necessary funds in financial markets. As we discussed in Chapter 18, this financing may take several forms: obtaining loans from banks, selling bonds to the public, or selling shares in future profits on the stock market. The neoclassical model assumes that if a firm is willing to pay the cost of capital, the financial markets will make the funds available.

Yet sometimes firms face **financing constraints**—limits on the amount they can raise in financial markets. Financing constraints can prevent firms from undertaking profitable investments. When a firm is unable to raise funds in financial markets, the amount it can spend on new capital goods is limited to the amount it is currently earning. Financing constraints influence the investment behavior of firms just as borrowing constraints influence the consumption behavior of households. Borrowing constraints cause households to determine their consumption based on current rather than permanent income; financing constraints cause firms to determine their investment based on their current cash flow rather than expected profitability.

To see the impact of financing constraints, consider the effect of a short recession on investment spending. A recession reduces employment, the rental price of capital, and profits. If firms expect the recession to be short-lived, however, they will want to continue investing, knowing that their investments will be profitable in the future. That is, a short recession will have only a small effect on Tobin's *q*. For firms that can raise funds in financial markets, the recession should have only a small effect on investment.

The opposite is true for firms that face financing constraints. The fall in current profits restricts the amount that these firms can spend on new capital goods and may prevent them from making profitable investments. Thus, financing constraints make investment more sensitive to current economic conditions.[12]

The extent to which financing constraints impede investment spending varies over time, depending on the health of the financial system, and this can in turn become a source of short-run fluctuations. As we discussed in Chapter 12, for example, during the Great Depression of the 1930s, many banks found themselves insolvent, as the value of their assets fell below the value of their liabilities. These banks were forced to suspend operations, making it more difficult for their customers to obtain financing for potential investment projects. Many economists believe the widespread bank failures during this period help explain the Depression's depth and persistence. Similarly, as we discussed in Chapters 12 and 18, the Great Recession of 2008–2009 came on the heels of a financial crisis.

The Bottom Line on Investment

The purpose of this section has been to examine the determinants of business fixed investment. We can reach three broad conclusions.

First, investment spending is inversely related to the real interest rate because a higher interest rate raises the cost of capital. Thus, the neoclassical model of investment justifies the investment function we have used throughout this book.

Second, various events can shift the investment function. An improvement in the available technology raises the marginal product of capital and raises investment. Various policies, such as changes in the corporate income tax, alter the incentives to invest and thus shift the investment function.

Third, investment will naturally be volatile over the business cycle because investment spending depends on the state of the economy as well as on the interest rate. In the neoclassical model of investment, higher employment raises the marginal product of capital and the incentive to invest. Higher output also raises firms' profits and, thereby, relaxes the financing constraints that some firms face. Our analysis predicts that an economic boom should stimulate investment, and a recession should depress it. This is exactly what we observe.

19-3 Conclusion: The Key Role of Expectations

Throughout our analysis of the microeconomic foundations of consumption and investment, one theme emerges: because households and firms are forward-looking, their expectations about the future influence the decisions they make today. People decide how much to consume by looking ahead to the income they expect to earn and to the standard of living they aspire to achieve. Business managers decide how much to invest by looking ahead to the profits that the new capital is likely to provide.

One corollary is that public policy influences consumption and investment not only through its direct impact but also by altering expectations. When deciding how much to respond to a tax hike or a tax cut, consumers anticipate whether the change is likely to be temporary or permanent. When making decisions about capital allocation, business managers consider the tax code they anticipate over the life of the investment. As a result, policymakers must take into account how their actions and words will influence the expectations of those making consumption and investment decisions.

In more advanced courses in macroeconomics, the modeling of expectations plays a large role. Some economists advocate the assumption of rational expectations, according to which decisionmakers optimally use available information, including information about public policy, when forecasting the future. Other economists suggest that deviations from conventional rationality, such as inattention and inertia, can help explain how people anticipate events. But there is broad consensus that expectations, however they are formed, are central to understanding economic behavior and the effects of policy.

EPILOGUE

Mankiw, *Macroeconomics*, 10e, © 2019 Worth Publishers

What We Know, What We Don’t

> If all the economists were laid end to end, they would not reach a conclusion.
>
> —George Bernard Shaw

> The theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions.
>
> —John Maynard Keynes

The first chapter of this book stated that the purpose of macroeconomics is to understand economic events and to improve economic policy. Having studied many of the most important models in the macroeconomist’s toolbox, we can now assess whether the field has achieved these goals.

Any fair assessment of macroeconomics today must admit that the science is incomplete. There are some principles that almost all macroeconomists accept and rely on when analyzing events or formulating policies. Yet many questions about the economy remain open to debate. In this epilogue, we review the central lessons of macroeconomics and the most pressing unresolved questions.

The Four Most Important Lessons of Macroeconomics

Let’s begin with four lessons that have recurred throughout this book and that most economists today endorse. Each lesson tells us how policy can influence a key economic variable—output, inflation, or unemployment—either in the long run or in the short run.

Lesson 1: In the long run, a country's capacity to produce goods and services determines the standard of living of its citizens.

Of all the statistics introduced in Chapter 2 and used throughout this book, the one that best captures economic well-being is GDP. Real GDP measures the economy's total output of goods and services and, therefore, a country's ability to satisfy the needs and desires of its citizens. Nations with higher GDP per person have more of almost everything—bigger homes, more cars, higher literacy, better health care, longer life expectancy, and more Internet connections. Perhaps the most important question in macroeconomics is what determines the level and the growth of GDP.

The models in Chapters 3, 8, and 9 identify the long-run determinants of GDP. In the long run, GDP depends on the factors of production—capital and labor—and on the technology for turning capital and labor into output. GDP grows when the factors of production increase or when the economy becomes better at transforming these inputs into goods and services.

This lesson has an important corollary: public policy can raise GDP in the long run only by improving the economy's productive capability. Policymakers can attempt to do this in many ways. Policies that raise national saving—by increasing either public or private saving—lead to a larger capital stock. Policies that raise the efficiency of labor—by supporting education or technological progress—lead to a more productive use of capital and labor. Policies that improve a nation's institutions—such as crackdowns on official corruption—promote both capital accumulation and the efficient use of scarce resources. By increasing the economy's output of goods and services, these policies enhance the standard of living.

Lesson 2: In the short run, aggregate demand influences the amount of goods and services that a country produces.

The economy's ability to *supply* goods and services is the sole determinant of GDP in the long run, but in the short run GDP also depends on the aggregate *demand* for goods and services. Aggregate demand is important because prices are sticky in the short run. The *IS–LM* model developed in Chapters 11 and 12, along with the open-economy Mundell–Fleming model in Chapter 13, shows what causes changes in aggregate demand and, therefore, short-run fluctuations in GDP.

Because aggregate demand influences output in the short run, all the variables that affect aggregate demand can influence economic fluctuations. Monetary policy, fiscal policy, and shocks to the money and goods markets are often responsible for year-to-year changes in output and employment. Because changes in aggregate demand are crucial to short-run fluctuations, policymakers monitor the economy closely. Before changing monetary or fiscal policy, they want to know whether the economy is booming or heading into a recession.

Lesson 3: In the long run, the rate of money growth determines the rate of inflation, but it does not affect the rate of unemployment.

In addition to GDP, inflation and unemployment are among the most closely watched measures of economic performance. Chapter 2 discussed how these two variables are measured, and subsequent chapters developed models to explain how they are determined.

The long-run analysis of Chapter 5 stresses that growth in the money supply is the ultimate determinant of inflation. That is, in the long run, a currency loses real value over time if and only if the central bank prints more and more of it. This lesson can explain the decade-to-decade variation in inflation observed in the United States, as well as the more dramatic hyperinflations that various countries have experienced from time to time.

We have also discussed many of the long-run effects of high inflation. In Chapter 5 we saw that, according to the Fisher effect, high inflation raises the nominal interest rate (so that the real interest rate remains unaffected). In Chapter 6 we saw that, high inflation causes the currency to depreciate in foreign exchange markets.

Warren Miller The New Yorker Collection/The Cartoon Bank

"And please let Jay Powell accept the things he cannot change, give him the courage to change the things he can, and the wisdom to know the difference."

The long-run determinants of unemployment are different. According to the classical dichotomy—the irrelevance of nominal variables in the determination of real variables—growth in the money supply does not affect unemployment in the long run. As we saw in Chapter 7, the natural rate of unemployment is determined by the rates of job separation and job finding, which in turn are determined by the process of job search and by the rigidity of the real wage.

Thus, we concluded that persistent inflation and persistent unemployment are unrelated problems. To combat inflation in the long run, policymakers must limit the growth in the money supply. To combat unemployment, they must improve the structure of labor markets. In the long run, there is no tradeoff between inflation and unemployment.

Lesson 4: In the short run, policymakers who control monetary and fiscal policy face a tradeoff between inflation and unemployment.

Although inflation and unemployment are not related in the long run, in the short run there is a tradeoff between these two variables, illustrated by the short-run Phillips curve. As we discussed in Chapter 14, policymakers can use monetary and fiscal policies to expand aggregate demand, lowering unemployment and raising inflation. Or they can use these policies to contract aggregate demand, raising unemployment and lowering inflation.

Policymakers face a fixed tradeoff between inflation and unemployment only in the short run. Over time, the short-run Phillips curve shifts for two reasons. First, supply shocks, such as changes in the price of oil, alter the short-run tradeoff; an adverse supply shock offers policymakers the difficult choice of higher inflation or higher unemployment. Second, when people change their expectations of inflation, the short-run tradeoff between inflation and unemployment shifts. The adjustment of expectations ensures that the tradeoff is temporary. That is, unemployment deviates from its natural rate only in the short run, and monetary policy has real effects only in the short run. In the long run, the classical model of Chapters 3 through 9 describes the world.

The Four Most Important Unresolved Questions of Macroeconomics

So far, we have been discussing the lessons about which most economists agree. Let’s now turn to four questions about which there is debate. Some disagreements concern the validity of alternative theories; others concern how theory should be applied to policy.

Question 1: How should policymakers try to promote growth in the economy’s natural level

of output?

Because the economy's natural level of output depends on capital, labor, and technology, any policy designed to raise output in the long run must aim to increase capital accumulation, improve the use of labor, or enhance the available technology. There is, however, no easy way to achieve these goals.

The Solow growth model of Chapters 8 and 9 shows that increasing the amount of capital requires raising the economy's rate of saving and investment. Therefore, many economists advocate policies to increase national saving. Yet the Solow model also shows that raising the capital stock requires a period of reduced consumption for current generations. Some argue that current generations should not be asked to make this sacrifice because technological progress will ensure that future generations are better off than current generations. (One waggish economist asked, "What has posterity ever done for me?") Even those who advocate increased saving and investment disagree about how to encourage saving and whether the investment should be in privately owned plants and equipment or in public infrastructure, such as roads and schools.

To improve the economy's use of its labor force, most policymakers would like to lower the natural rate of unemployment. As we discussed in Chapter 7, the differences in unemployment that we observe across countries, as well as the changes we observe over time, suggest that the natural rate is not an immutable constant but depends on a nation's policies and institutions. Yet reducing unemployment is fraught with perils. The natural rate of unemployment could be reduced by decreasing unemployment-insurance benefits (and thus increasing the search effort of the unemployed) or by decreasing the minimum wage (and thus bringing wages closer to equilibrium levels). Yet these policies would also hurt some members of society most in need and, therefore, do not command a consensus among economists. During the Great Recession of 2008–2009, the U.S. Congress temporarily extended eligibility for unemployment insurance to an unprecedented 99 weeks, generating a debate about whether this was an appropriate response to extraordinary circumstances or an overreaction.

In many countries, the natural level of output is depressed by a lack of institutions that people in developed nations take for granted. U.S. citizens today do not worry about revolutions, coups, or civil wars. They generally trust the police and the court system to respect the laws, maintain order, protect property rights, and enforce private contracts. In nations without such institutions, people face the wrong incentives: if creating something of value is a less reliable path to riches than is stealing from a neighbor, an economy is unlikely to prosper. All economists agree that establishing the right institutions is necessary for increasing growth in the world's poor nations, but transforming a nation's institutions requires overcoming difficult political hurdles.

Promoting technological progress is, according to some economists, the most important objective for public policy. The Solow growth model shows that only technological progress can yield persistent growth in living standards. Despite much work on theories of endogenous growth, which highlight the societal decisions that influence technological change, economists cannot offer a reliable recipe to ensure rapid advances in technology. They continue to debate the extent to which the government should actively promote the development and expansion of particular industries and technologies.

Question 2: Should policymakers try to stabilize the economy? If so, how?

The model of aggregate supply and aggregate demand developed in Chapters 10 through 15 shows how shocks to the economy cause economic fluctuations and how monetary and fiscal policy can influence these fluctuations. Some economists believe that policymakers should use this analysis to stabilize the economy. They believe that monetary and fiscal policy should try to offset shocks to keep output and employment near their natural levels.

Yet, as we discussed in Chapter 16, others are skeptical about our ability to stabilize the economy, citing the long and variable lags inherent in policymaking, the poor record of forecasting, and our still-limited understanding of how the economy works. These economists conclude that it is best for policy to be more passive. In addition, many economists believe that all too often policymakers are politically opportunistic or tempted to follow time-inconsistent policies. They conclude that policymakers should not have discretion over monetary and fiscal policy but should instead be committed to a policy rule. Or, at the very least, their discretion should be constrained, as is the case when central banks adopt a policy of inflation targeting.

There is also debate among economists about which macroeconomic tools are best suited for stabilization. Typically, monetary policy is the first line of defense against the business cycle. In the Great Recession of 2008–2009, however, the Fed cut interest rates to their lower bound of zero, and the focus turned to fiscal policy. Economists often disagree about the extent to which fiscal policy should be used to stimulate the economy in downturns and whether tax cuts or spending increases are the preferred fiscal tool.

A related question is whether the benefits of stabilization—assuming it could be achieved—would be large or small. Without a change in the natural rate of unemployment, stabilization policy can only reduce the magnitude of fluctuations around the natural rate. Thus, successful stabilization policy would eliminate booms as well as recessions. Some economists have suggested that the average gain from stabilization would be small.

Finally, in the aftermath of the financial crisis and Great Recession of 2008–2009, economists questioned whether the economy could be stabilized by avoiding such shocks in the future. As we discussed in Chapter 18, problems in the financial system can lead to problems throughout the economy. Indeed, over the course of history, financial crises have caused some of the deepest downturns. Unfortunately, it is not clear how best to prevent such crises.

One point of debate centers on how monetary policy should respond to speculative bubbles in asset prices. Some economists argue that central banks should monitor asset markets and try to prevent speculative bubbles. This might mean raising interest rates earlier than otherwise to deflate bubbles as they begin. Other economists believe that monetary policymakers are no better than market participants at telling when a rise in asset prices reflects an irrational speculative bubble rather than a rational evaluation of changing fundamentals. Moreover, they argue, the tools of monetary policy are too crude to prick bubbles, and trying to do so could distract central banks from their main goals of low inflation and stable employment.

Another point of debate concerns regulation. Some economists argue that more vigilant regulation of financial institutions can reduce reckless risk-taking and the likelihood of financial crises. Others believe that financial regulation is hard to execute, easy to circumvent, and liable to give the public a false hope that the financial system is safer than it really is. In addition, they argue that excessive regulation could divert the financial system from performing its job of efficiently allocating capital and risk, thereby impeding long-run growth.

Question 3: How costly is inflation, and how costly is reducing inflation?

Whenever prices are rising, policymakers face the question of whether to pursue policies to reduce inflation. To make this decision, they must compare the cost of allowing inflation to continue at its current rate to the cost of reducing it. Yet economists cannot offer accurate estimates of either cost.

The cost of inflation is a topic on which economists and laypeople often disagree. When inflation reached 10 percent per year in the late 1970s, polls showed that the public viewed inflation as a major problem. Yet, as we saw in Chapter 5, when economists try to identify the social costs of inflation, they can point only to a few costs, including shoeleather costs, menu costs, and the costs of a nonindexed tax system. These costs are large during hyperinflations, but they seem minor at the moderate rates of inflation experienced in most major economies. Some economists believe that the public confuses inflation with other problems that coincide with inflation. For example, as growth in productivity and real wages slowed in the 1970s, some laypeople might have viewed inflation as the cause of the slowdown in real wages. Yet economists may be mistaken: perhaps inflation is very costly, and we have yet to figure out why.

It is also possible that some inflation is desirable. If workers resist cuts in nominal wages, then inflation makes it easier for real wages to fall when necessary to equilibrate the supply and demand for labor. That is, inflation may “grease the wheels” of labor markets. In addition, higher inflation raises the nominal interest rate through the Fisher effect, and a higher nominal interest rate gives the central bank more room to cut interest rates when necessary to stimulate the economy. In other words, higher inflation makes it less likely that the central bank will hit the zero lower bound on nominal interest rates, reducing the risk of a liquidity trap. Some economists use these arguments to suggest that the Fed aim for 4 percent inflation instead of its current 2 percent target.

The cost of reducing inflation is a topic on which economists often disagree among themselves. As we discussed in Chapter 14, the standard view—as described by the short-run Phillips curve—is that reducing inflation requires a period of low output and high unemployment. According to this view, the cost of reducing inflation is measured by the sacrifice ratio, the number of percentage points of a year's GDP that must be forgone to reduce inflation by 1 percentage point. But some economists think that reducing inflation can be less costly than estimates of the sacrifice ratio indicate. According to the rational-expectations approach discussed in Chapter 14, if a disinflationary policy is announced in advance and is credible, people will adjust their expectations quickly, so the disinflation need not cause a recession.

Other economists believe that the cost of reducing inflation is larger than indicated by estimates of the sacrifice ratio. Theories of hysteresis discussed in Chapter 14 suggest that a recession caused by disinflationary policy could raise the natural rate of unemployment. If so, the cost of reducing inflation is not a temporary recession but a persistently higher level of unemployment.

Because the costs of inflation and disinflation remain open to debate, economists sometimes offer conflicting advice to policymakers. Perhaps with further research, we can reach a consensus on the optimal rate of inflation and the best way to achieve it.

Question 4: How big a problem are government budget deficits?

Government debt is a perennial topic of debate, particularly in recent years. During the Great Recession of 2008–2009, the U.S. budget deficit increased to $1.4 trillion, or about 10 percent of GDP, a level not seen since World War II. Even more troubling is the long-term fiscal picture. Many economists believe that the budget deficit will be hard to control as the large baby-boom generation reaches retirement age and starts drawing on the benefits that the government provides to the elderly.

Most economists take the traditional view of government debt. According to this view, when the government runs a budget deficit and issues debt, it reduces national saving, leading to lower investment and a trade deficit. In the long run, it leads to a smaller steady-state capital stock and a larger foreign debt. Those who hold the traditional view conclude that government debt places a burden on future generations.

Yet, as we discussed in Chapter 17, some economists are skeptical of this assessment. Advocates of the

Ricardian view of government debt stress that a budget deficit represents a substitution of future taxes for current taxes. As long as consumers are forward-looking, as many theories of consumption presented in Chapter 19 assume, they will save today to meet their or their children's future tax liability. These economists believe that the level of government debt has a minor effect on the economy. They maintain that the government's spending decisions matter, but whether that spending is financed by taxation or by selling bonds is of secondary importance.

Still other economists assert that conventional measures of fiscal policy are too flawed to be of much use. Although the government's choices regarding taxes and spending have great influence on the welfare of different generations, not all of these choices are reflected in measures of the government debt. The level of Social Security benefits and taxes, for instance, affects the welfare of the elder beneficiaries versus that of the working-age taxpayers, but the size of the budget deficit does not reflect this policy choice. According to some economists, we should stop focusing on the budget deficit and concentrate instead on the broader generational impacts of fiscal policy.

Recent events have focused renewed attention on the possibility of government default. In the eighteenth century, Alexander Hamilton argued successfully that the U.S. federal government should always honor its debts. But in recent years Greece and several other European nations have struggled to do so. In August 2011, Standard & Poor's reduced its credit rating on U.S. bonds below the top AAA level, where it still remains, suggesting that Hamilton's rule might someday be violated even in the United States. As the U.S. political system wrestles with budget deficits, both economists and the public are divided about what should be done to return fiscal policy to a sustainable path. Reasonable people disagree about how much of the fiscal adjustment should come from higher tax revenue and how much should come from reduced government spending.

Conclusion

Economists and policymakers must deal with ambiguity. The current state of macroeconomics offers many insights but also leaves many questions open. The challenge for economists is to answer these questions and expand our knowledge. The challenge for policymakers is to use the knowledge we have to improve economic performance. Both challenges are formidable, but neither is insuperable.

Glossary

Accommodating policy:
A policy that yields to the effect of a shock and thereby prevents the shock from being disruptive; for example, a policy that raises aggregate demand in response to an adverse supply shock, sustaining the effect of the shock on prices and keeping output at its natural level.

Accounting profit:
The amount of revenue remaining for the owners of a firm after all the factors of production except capital have been compensated. (Cf. economic profit, profit.)

Acyclical:
Moving in no consistent direction over the business cycle. (Cf. countercyclical, procyclical.)

Adaptive expectations:
An approach that assumes that people form their expectation of a variable based on recently observed values of the variable. (Cf. rational expectations.)

Adverse selection:
An unfavorable sorting of individuals by their own choices; for example, in efficiency-wage theory, when a wage cut induces good workers to quit and bad workers to remain with the firm.

Aggregate:
Total for the whole economy.

Aggregate demand:
The negative relationship between the price level and the aggregate quantity of output demanded that arises from the interaction between the goods market and the money market.

Aggregate supply:
The relationship between the price level and the aggregate quantity of output firms produce.

Animal spirits:
Exogenous and perhaps self-fulfilling waves of optimism and pessimism about the state of the economy that, according to some economists, influence the level of investment.

Appreciation:
A rise in the value of a currency relative to other currencies in the market for foreign exchange. (Cf. depreciation.)

Arbitrage:
The act of buying an item in one market and selling it at a higher price in another market in order to profit from the price differential in the two markets.

Asymmetric information:

A situation in which one party in an economic transaction has some relevant information not available to the other party.

Automatic stabilizer:
A policy that reduces the amplitude of economic fluctuations without regular and deliberate changes in economic policy; for example, an income tax system that automatically reduces taxes when income falls.

Average propensity to consume (*APC*):
The ratio of consumption to income (C/Y).

Balance sheet:
An accounting statement that shows assets and liabilities.

Balanced budget:
A budget in which receipts equal expenditures.

Balanced growth:
The condition under which many economic variables, such as income per person, capital per person, and the real wage, all grow at the same rate.

Balanced trade:
A situation in which the value of imports equals the value of exports, so net exports equal zero.

Bank capital:
The resources the bank owners have put into the institution.

Bond:
A document representing an interest-bearing debt of the issuer, usually a corporation or the government.

Borrowing constraint:
A restriction on the amount a person can borrow from financial institutions, limiting that person’s ability to spend his or her future income today; also called a liquidity constraint.

Budget deficit:
A shortfall of receipts from expenditure.

Budget surplus:
An excess of receipts over expenditure.

Business cycle:
Economy-wide fluctuations in output, incomes, and employment.

Capital:
1. The stock of equipment and structures used in production. 2. The funds to finance the accumulation of equipment and structures.

Capital budgeting:
An accounting procedure that measures both assets and liabilities.

Capital requirement:
A minimum amount of bank capital mandated by regulators.

Central bank:
The institution responsible for the conduct of monetary policy, such as the Federal Reserve in the United States.

Classical dichotomy:
The theoretical separation of real and nominal variables in the classical model, which implies that nominal variables do not influence real variables. (Cf. neutrality of money.)

Classical model:
A model of the economy derived from the ideas of the classical, or pre-Keynesian, economists; a model based on the assumptions that wages and prices adjust to clear markets and that monetary policy does not influence real variables. (Cf. Keynesian model.)

Closed economy:
An economy that does not engage in international trade. (Cf. open economy.)

Cobb–Douglas production function:
A production function of the form $F(K, L) = AK^{\alpha}L^{1-\alpha}$, where K is capital, L is labor, and A and α are parameters.

Commodity money:
Money that is intrinsically useful and would be valued even if it did not serve as money. (Cf. fiat money, money.)

Competition:
A situation in which there are many individuals or firms, so that the actions of any one of them do not influence market prices.

Conditional convergence:
The tendency of economies with different initial levels of income but similar economic policies and institutions to become more similar in income over time.

Constant returns to scale:
A property of a production function whereby a proportionate increase in all factors of production leads to an increase in output of the same proportion.

Consumer price index (CPI):
A measure of the overall level of prices that shows the cost of a fixed basket of consumer goods relative to the cost of the same basket in a base year.

Consumption:
Goods and services purchased by consumers.

Consumption function:
A relationship showing the determinants of consumption; for example, a relationship between consumption

and disposable income, $C = C(Y - T)$.

Contractionary policy:
Policy that reduces aggregate demand, real income, and employment. (Cf. expansionary policy.)

Convergence:
The tendency of economies with different initial levels of income to become more similar in income over time.

Corporate income tax:
The tax levied on the accounting profit of corporations.

Cost of capital:
The amount forgone by holding a unit of capital for one period, including interest, depreciation, and the gain or loss from the change in the price of capital.

Cost-push inflation:
Inflation resulting from shocks to aggregate supply. (Cf. demand-pull inflation.)

Countercyclical:
Moving in the opposite direction from output, incomes, and employment over the business cycle; rising during recessions and falling during recoveries. (Cf. acyclical, procyclical.)

CPI:
See consumer price index.

Creative destruction:
The process whereby entrepreneurs introduce innovations that render some incumbent producers unprofitable while promoting overall economic growth.

Credit crunch:
A change in conditions at financial institutions that makes it hard for potential borrowers to obtain loans.

Crowding out:
The reduction in investment that results when expansionary fiscal policy raises the interest rate.

Currency:
The sum of outstanding paper money and coins.

Currency board:
A fixed exchange rate system under which a central bank backs all of the nation's currency with the currency of another country.

Currency–deposit ratio:
The ratio of the amount of currency that people choose to hold to the amount of demand deposits they hold at banks.

Cyclical unemployment:

The unemployment associated with short-run economic fluctuations; the deviation of the unemployment rate from the natural rate.

Cyclically adjusted budget deficit:
The budget deficit adjusted for the influence of the business cycle on government spending and tax revenue; the budget deficit that would occur if the economy's production and employment were at their natural levels. Also called full-employment budget deficit.

Debt-deflation theory:
A theory according to which an unexpected fall in the price level redistributes real wealth from debtors to creditors and, therefore, reduces total spending in the economy.

Debt finance:
The process of obtaining funds for a business by borrowing, such as through the bond market.

Deflation:
A decrease in the overall level of prices. (Cf. disinflation, inflation.)

Deflator:
See GDP deflator or PCE deflator.

Demand deposits:
Assets that are held in banks and can be used on demand to make transactions, such as checking accounts.

Demand-pull inflation:
Inflation resulting from shocks to aggregate demand. (Cf. cost-push inflation.)

Demand shocks:
Exogenous events that shift the aggregate demand curve.

Depreciation:
1. The reduction in the capital stock that occurs over time because of aging and use. 2. A fall in the value of a currency relative to other currencies in the market for foreign exchange. (Cf. appreciation.)

Depression:
A very severe recession.

Devaluation:
An action by the central bank to decrease the value of a currency under a system of fixed exchange rates. (Cf. revaluation.)

Diminishing marginal product:
A characteristic of a production function whereby the marginal product of a factor falls as the amount of the factor increases while all other factors are held constant.

Discount rate:
The interest rate that the Fed charges when it makes loans to banks.

Discouraged workers:
Individuals who have left the labor force because they believe there is little hope of finding a job.

Disinflation:
A reduction in the rate at which prices are rising. (Cf. deflation, inflation.)

Disposable income:
Income remaining after the payment of taxes.

Diversification:
Reduction of risk by holding assets with imperfectly correlated returns.

Dollarization:
The adoption of the U.S. dollar as the currency in another country.

Double coincidence of wants:
A situation in which two individuals each have precisely the good that the other wants.

Economic profit:
The amount of revenue remaining for the owners of a firm after all the factors of production have been compensated. (Cf. accounting profit, profit.)

Efficiency of labor:
A variable in the Solow growth model that measures the health, education, skills, and knowledge of the labor force.

Efficiency units of labor:
A measure of the labor force that incorporates both the number of workers and the efficiency of each worker.

Efficiency-wage theories:
Theories of real-wage rigidity and unemployment according to which firms raise labor productivity and profits by keeping real wages above the equilibrium level.

Efficient markets hypothesis:
The theory that asset prices reflect all publicly available information about the value of an asset.

Elasticity:
The percentage change in a variable caused by a 1-percent change in another variable.

Endogenous growth theory:
Models of economic growth that try to explain the rate of technological change.

Endogenous variable:
A variable that is explained by a particular model; a variable whose value is determined by the model's solution. (Cf. exogenous variable.)

Equilibrium:

A state of balance between opposing forces, such as the balance of supply and demand in a market.

Equity finance:
The process of obtaining funds for a business by issuing ownership shares, such as through the stock market.

Euler's theorem:
The mathematical result economists use to show that economic profit must be zero if the production function has constant returns to scale and if factors are paid their marginal products.

***Ex ante* real interest rate**:
The real interest rate anticipated when a loan is made; the nominal interest rate minus expected inflation. (Cf. *ex post* real interest rate.)

***Ex post* real interest rate**:
The real interest rate actually realized; the nominal interest rate minus actual inflation. (Cf. *ex ante* real interest rate.)

Excess reserves:
Reserves held by banks above the amount mandated by reserve requirements.

Exchange rate:
The rate at which a country makes exchanges in world markets. (Cf. nominal exchange rate, real exchange rate.)

Exogenous variable:
A variable that a particular model takes as given; a variable whose value is independent of the model's solution. (Cf. endogenous variable.)

Expansionary policy:
Policy that raises aggregate demand, real income, and employment. (Cf. contractionary policy.)

Exports:
Goods and services sold to other countries.

Factor of production:
An input used to produce goods and services; for example, capital or labor.

Factor price:
The amount paid for one unit of a factor of production.

Factor share:
The proportion of total income being paid to a factor of production.

Federal funds rate:
The overnight interest rate at which banks lend to one another.

Federal Reserve (the Fed):

The central bank of the United States.

Fiat money:
Money that is not intrinsically useful and is valued only because it is used as money. (Cf. commodity money, money.)

Financial crisis:
A major disruption in the financial system that impedes the economy’s ability to intermediate between those who want to save and those who want to borrow and invest.

Financial intermediaries:
Institutions that facilitate the matching of savers and borrowers, such as banks.

Financial intermediation:
The process by which resources are allocated from those individuals who wish to save some of their income for future consumption to those individuals and firms who wish to borrow to buy investment goods for future production.

Financial markets:
Markets through which savers can directly provide resources to borrowers, such as the stock market and bond market.

Financial system:
The set of institutions through which the resources of those who want to save are allocated to those who want to borrow.

Financing constraint:
A limit on the quantity of funds a firm can raise—such as through borrowing—in order to buy capital.

Fire sale:
The precipitous fall in the price of assets that takes place when financial institutions must sell their assets quickly in the midst of a crisis.

Fiscal policy:
The government’s choice regarding levels of spending and taxation.

Fisher effect:
The one-for-one influence of expected inflation on the nominal interest rate.

Fisher equation:
The equation stating that the nominal interest rate is the sum of the real interest rate and expected inflation ($i = r + E\pi$).

Fixed exchange rate:
An exchange rate that is set by the central bank’s willingness to buy and sell the domestic currency for foreign currencies at a predetermined price. (Cf. floating exchange rate.)

Flexible prices:

Prices that adjust quickly to equilibrate supply and demand. (Cf. sticky prices.)

Floating exchange rate:
An exchange rate that the central bank allows to change in response to changing economic conditions and economic policies. (Cf. fixed exchange rate.)

Flow:
A variable measured as a quantity per unit of time. (Cf. stock.)

Fractional-reserve banking:
A system in which banks keep only some of their deposits on reserve. (Cf. 100-percent-reserve banking.)

Frictional unemployment:
The unemployment that results because it takes time for workers to search for the jobs that best suit their skills and tastes. (Cf. structural unemployment.)

Full-employment budget deficit:
See cyclically adjusted budget deficit.

GDP:
See gross domestic product.

GDP deflator:
The ratio of nominal GDP to real GDP; a measure of the overall level of prices that shows the cost of the currently produced basket of goods relative to the cost of that basket in a base year.

General equilibrium:
The simultaneous equilibrium of all the markets in the economy.

GNP:
See gross national product.

Gold standard:
A monetary system in which gold serves as money or in which all money is convertible into gold at a fixed rate.

Golden Rule:
The saving rate in the Solow growth model that leads to the steady state in which consumption per worker (or consumption per efficiency unit of labor) is maximized.

Government purchases:
Goods and services bought by the government. (Cf. transfer payments.)

Government-purchases multiplier:
The change in aggregate income resulting from a one-dollar change in government purchases.

Gross domestic product (GDP):
The total income earned domestically, including the income earned by foreign-owned factors of production;

the total expenditure on domestically produced goods and services.

Gross national product (GNP):
The total income of all residents of a nation, including the income from factors of production used abroad; the total expenditure on the nation’s output of goods and services.

High-powered money:
The sum of currency and bank reserves; also called the monetary base.

Human capital:
The accumulation of investments in people, such as education.

Hyperinflation:
Extremely high inflation.

Hysteresis:
The long-lasting influence of history, such as on the natural rate of unemployment.

Imperfect-information model:
The model of aggregate supply emphasizing that individuals do not always know the overall price level because they cannot observe the prices of all goods and services in the economy.

Import quota:
A legal limit on the amount of a good that can be imported.

Imports:
Goods and services bought from other countries.

Impossible trinity:
The fact that a nation cannot simultaneously have free capital flows, a fixed exchange rate, and independent monetary policy. Sometimes called the trilemma of international finance.

Imputed value:
An estimate of the value of a good or service that is not sold in the marketplace and therefore does not have a market price.

Income velocity of money:
The ratio of national income, as measured by GDP, to the money supply.

Index of leading indicators:
See leading indicators.

Inflation:
An increase in the overall level of prices. (Cf. deflation, disinflation.)

Inflation rate:
A measure of how fast prices are rising.

Inflation targeting:

A monetary policy under which the central bank announces a specific target, or target range, for the inflation rate.

Inflation tax:
The revenue raised by the government through the creation of money; also called seigniorage.

Inside lag:
The time between a shock hitting the economy and the policy action taken to respond to the shock. (Cf. outside lag.)

Insiders:
Workers who are already employed and therefore have an influence on wage bargaining. (Cf. outsiders.)

Interest on reserves:
The central bank's policy of paying banks an interest rate for the deposits that they hold as reserves.

Interest rate:
The market price at which resources are transferred between the present and the future; the return to saving and the cost of borrowing.

Intermediation:
See financial intermediation.

Investment:
Goods purchased by individuals and firms to add to their stock of capital.

Investment tax credit:
A provision of the corporate income tax that reduces a firm's tax when it buys new capital goods.

***IS* curve**:
The negative relationship between the interest rate and the level of income that arises in the market for goods and services. (Cf. *IS–LM* model, *LM* curve.)

***IS–LM* model**:
A model of aggregate demand that shows what determines aggregate income for a given price level by analyzing the interaction between the goods market and the money market. (Cf. *IS* curve, *LM* curve.)

Keynesian cross:
A simple model of income determination, based on the ideas in Keynes's *General Theory*, which shows how changes in spending can have a multiplied effect on aggregate income.

Keynesian model:
A model derived from the ideas of Keynes's *General Theory*; a model based on the assumptions that wages and prices do not adjust to clear markets and that aggregate demand determines the economy's output and employment. (Cf. classical model.)

Labor-augmenting technological progress:
Advances in productive capability that raise the efficiency of labor.

Labor force:
Those in the population who have a job or are looking for a job.

Labor-force participation rate:
The percentage of the adult population in the labor force.

Labor hoarding:
The phenomenon of firms employing workers whom they do not need when the demand for their products is low so that they will still have these workers when demand recovers.

Large open economy:
An open economy that can influence its domestic interest rate; an economy that, by virtue of its size, can have a substantial impact on world markets and, in particular, on the world interest rate. (Cf. small open economy.)

Laspeyres price index:
A measure of the level of prices based on a fixed basket of goods. (Cf. Paasche price index.)

Leading indicators:
Economic variables that fluctuate in advance of the economy's output and thus signal the direction of economic fluctuations.

Lender of last resort:
The role a central bank plays when it lends to financial institutions in the midst of a liquidity crisis.

Leverage:
The use of borrowed money to supplement existing funds for purposes of investment.

Life-cycle hypothesis:
The theory of consumption that emphasizes the role of saving and borrowing as transferring resources from those times in life when income is high to those times in life when income is low, such as from working years to retirement.

Liquid:
Readily convertible into the medium of exchange; easily used to make transactions.

Liquidity constraint:
A restriction on the amount a person can borrow from a financial institution, which limits the person's ability to spend his future income today; also called a borrowing constraint.

Liquidity crisis:
A situation in which a solvent bank does not have sufficient cash on hand to satisfy the withdrawal demands of depositors.

Liquidity trap:
A situation in which the nominal interest rate has fallen to its lower bound of zero, calling into question the efficacy of monetary policy to further stimulate the economy.

Liquidity-preference theory:
See theory of liquidity preference.

***LM* curve**:
The positive relationship between the interest rate and the level of income (while holding the price level fixed) that arises in the market for real money balances. (Cf. *IS*–*LM* model, *IS* curve.)

Loanable funds:
The flow of resources available to finance capital accumulation.

Lucas critique:
The argument that traditional policy analysis does not adequately take into account the impact of policy changes on people's expectations.

***M*1, *M*2**:
Various measures of the stock of money, where larger numbers signify a broader definition of money.

Macroeconometric model:
A model that uses data and statistical techniques to describe the economy quantitatively rather than just qualitatively.

Macroeconomics:
The study of the economy as a whole. (Cf. microeconomics.)

Macroprudential regulation:
Regulation of financial institutions that focuses on system-wide risks.

Marginal product of capital (*MPK*):
The amount of extra output produced when the capital input is increased by one unit.

Marginal product of labor (*MPL*):
The amount of extra output produced when the labor input is increased by one unit.

Marginal propensity to consume (*MPC*):
The increase in consumption resulting from a one-dollar increase in disposable income.

Market-clearing model:
A model that assumes that prices freely adjust to equilibrate supply and demand.

Medium of exchange:
An item widely accepted in transactions for goods and services; one of the functions of money. (Cf. store of value, unit of account.)

Menu cost:
The cost of changing a price.

Microeconomics:
The study of individual markets and decisionmakers. (Cf. macroeconomics.)

Microprudential regulation:
Regulation of financial institutions that focuses on the risks facing individual institutions.

Model:
A simplified representation of reality, often using diagrams or equations, that shows how variables interact.

Monetarism:
The doctrine according to which changes in the money supply are the primary cause of economic fluctuations, implying that a stable money supply would lead to a stable economy.

Monetary base:
The sum of currency and bank reserves; also called high-powered money.

Monetary neutrality:
See neutrality of money.

Monetary policy:
The central bank's choice regarding the supply of money.

Monetary transmission mechanism:
The process by which changes in the money supply influence the amount that households and firms wish to spend on goods and services.

Monetary union:
A group of economies that have decided to share a common currency and thus a common monetary policy.

Money:
The stock of assets used for transactions. (Cf. commodity money, fiat money.)

Money demand function:
A function showing the determinants of the demand for real money balances; for example, $(M/P)^d = L(i, Y)$.

Money multiplier:
The increase in the money supply resulting from a one-dollar increase in the monetary base.

Money supply:
The amount of money available, usually as determined by the central bank and the banking system.

Moral hazard:
The possibility of dishonest or otherwise inappropriate behavior in situations in which behavior is imperfectly monitored; for example, in efficiency-wage theory, the possibility that low-wage workers may shirk their responsibilities and risk getting caught and fired.

Multiplier:
See government-purchases multiplier, money multiplier, or tax multiplier.

Mundell–Fleming model:
The *IS–LM* model for a small open economy.

Mutual fund:
A financial intermediary that holds a diversified portfolio of stock or bonds.

NAIRU:
Non-accelerating inflation rate of unemployment.

National income accounting:
The accounting system that measures GDP and many other related statistics.

National income accounts identity:
The equation showing that GDP is the sum of consumption, investment, government purchases, and net exports.

National saving:
A nation’s income minus consumption and government purchases; the sum of private and public saving.

Natural rate of unemployment:
The steady-state rate of unemployment; the rate of unemployment toward which the economy gravitates in the long run.

Natural-rate hypothesis:
The premise that fluctuations in aggregate demand influence output, employment, and unemployment only in the short run and that in the long run these variables return to the levels implied by the classical model.

Neoclassical model of investment:
The theory according to which investment depends on the deviation of the marginal product of capital from the cost of capital.

Net capital outflow:
The net flow of funds being invested abroad; domestic saving minus domestic investment; also called net foreign investment.

Net exports:
Exports minus imports.

Net foreign investment:
See net capital outflow.

Net investment:
The amount of investment after the replacement of depreciated capital; the change in the capital stock.

Neutrality of money:
The property that a change in the money supply does not influence real variables. (Cf. classical dichotomy.)

Nominal:
Measured in current dollars; not adjusted for inflation. (Cf. real.)

Nominal exchange rate:
The rate at which one country's currency trades for another country's currency. (Cf. exchange rate, real exchange rate.)

Nominal interest rate:
The return to saving and the cost of borrowing without adjustment for inflation. (Cf. real interest rate.)

Okun's law:
The negative relationship between unemployment and real GDP, according to which a decrease in unemployment of 1 percentage point is associated with additional growth in real GDP of approximately 2 percent.

100-percent-reserve banking:
A system in which banks keep all deposits on reserve. (Cf. fractional reserve banking.)

Open economy:
An economy in which people can freely engage in international trade in goods and capital. (Cf. closed economy.)

Open-market operations:
The purchase or sale of government bonds by the central bank for the purpose of increasing or decreasing the money supply.

Optimize:
To achieve the best possible outcome subject to a set of constraints.

Outside lag:
The time between a policy action and its influence on the economy. (Cf. inside lag.)

Outsiders:
Workers who are not employed and therefore have no influence on wage bargaining. (Cf. insiders.)

Paasche price index:
A measure of the level of prices based on a changing basket of goods. (Cf. Laspeyres price index.)

PCE deflator:
The ratio of nominal personal consumption expenditure to real personal consumption expenditure; a measure of the overall level of prices that shows the cost of the currently consumed basket of goods relative to the cost of that basket in a base year.

Permanent income:
Income that people expect to persist into the future; normal income. (Cf. transitory income.)

Permanent-income hypothesis:
The theory of consumption according to which people choose consumption based on their permanent

income and use saving and borrowing to smooth consumption in response to transitory variations in income.

Phillips curve:
A negative relationship between inflation and unemployment; in its modern form, a relationship among inflation, cyclical unemployment, expected inflation, and supply shocks, derived from the short-run aggregate supply curve.

Pigou effect:
The increase in consumer spending that results when a fall in the price level raises real money balances and, thereby, consumers' wealth.

Political business cycle:
The fluctuations in output and employment resulting from the manipulation of the economy for electoral gain.

Predetermined variable:
A variable whose value was fixed in a previous period of time.

Present value:
The amount today that is equivalent to an amount to be received in the future, taking into account the interest that could be earned over the interval of time.

Private saving:
Disposable income minus consumption.

Procyclical:
Moving in the same direction as output, incomes, and employment over the business cycle; falling during recessions and rising during recoveries. (Cf. acyclical, countercyclical.)

Production function:
The mathematical relationship showing how the quantities of the factors of production determine the quantity of goods and services produced; for example, $Y = F(K, L)$.

Profit:
The income of firm owners; firm revenue minus firm costs. (Cf. accounting profit, economic profit.)

Public saving:
Government receipts minus government spending; the budget surplus.

Purchasing-power parity:
The doctrine according to which goods must sell for the same price in every country, implying that the nominal exchange rate reflects differences in price levels.

***q* theory of investment**:
The theory according to which expenditure on capital goods depends on the ratio of the market value of installed capital to its replacement cost.

Quantity equation:
The identity stating that the product of the money supply and the velocity of money equals nominal expenditure ($MV = PY$); coupled with the assumption of stable velocity, an explanation of nominal expenditure called the quantity theory of money.

Quantity theory of money:
The doctrine emphasizing that changes in the quantity of money lead to changes in nominal expenditure.

Quota:
See import quota.

Random variable:
A variable whose value is determined by chance.

Random walk:
The path followed by a variable whose changes over time are unpredictable.

Rational expectations:
An approach that assumes that people optimally use all available information—including information about current and prospective policies—to forecast the future. (Cf. adaptive expectations.)

Real:
Measured in constant dollars; adjusted for inflation. (Cf. nominal.)

Real business cycle theory:
The theory according to which economic fluctuations can be explained by real changes in the economy (such as changes in technology) and without any role for nominal variables (such as the money supply).

Real cost of capital:
The cost of capital adjusted for the overall price level.

Real exchange rate:
The rate at which one country's goods trade for another country's goods. (Cf. exchange rate, nominal exchange rate.)

Real interest rate:
The return to saving and the cost of borrowing after adjustment for inflation. (Cf. nominal interest rate.)

Real money balances:
The quantity of money expressed in terms of the quantity of goods and services it can buy; the quantity of money divided by the price level (M/P).

Recession:
A sustained period of falling real income.

Rental price of capital:
The amount paid to rent one unit of capital.

Reserve–deposit ratio:
The ratio of the amount of reserves banks choose to hold to the amount of demand deposits they have.

Reserve requirements:
Regulations imposed on banks by the central bank that specify a minimum reserve–deposit ratio.

Reserves:
The money that banks have received from depositors but have not used to make loans.

Revaluation:
An action undertaken by the central bank to raise the value of a currency under a system of fixed exchange rates. (Cf. devaluation.)

Ricardian equivalence:
The theory according to which forward-looking consumers fully anticipate the future taxes implied by government debt, so that government borrowing today coupled with a tax increase in the future to repay the debt has the same effect on the economy as a tax increase today.

Risk aversion:
A dislike of uncertainty.

Sacrifice ratio:
The number of percentage points of a year's real GDP that must be forgone to reduce inflation by 1 percentage point.

Saving:
See national saving, private saving, and public saving.

Seasonal adjustment:
The removal of the regular fluctuations in an economic variable that occur as a function of the time of year.

Sectoral shift:
A change in the composition of demand among industries or regions.

Seigniorage:
The revenue raised by the government through the creation of money; also called the inflation tax.

Shadow banks:
Financial institutions that (like banks) are at the center of financial intermediation but (unlike banks) do not take in deposits insured by the FDIC.

Shock:
An exogenous change in an economic relationship, such as the aggregate demand or aggregate supply curve.

Shoeleather cost:
The cost of inflation from reducing real money balances, such as the inconvenience of needing to make more frequent trips to the bank.

Small open economy:
An open economy that takes its interest rate as given by world financial markets; an economy that, by virtue of its size, has a negligible impact on world markets and, in particular, on the world interest rate. (Cf. large open economy.)

Solow growth model:
A model showing how saving, population growth, and technological progress determine the level of and growth in the standard of living.

Solow residual:
The growth in total factor productivity, measured as the percentage change in output minus the percentage change in inputs, where the inputs are weighted by their factor shares. (Cf. total factor productivity.)

Speculative attack:
The massive selling of a country's currency, often because of a change in investors' perceptions, that renders a fixed exchange rate untenable.

Speculative bubble:
A rise in the price of an asset above its fundamental value.

Stabilization policy:
Public policy aimed at reducing the severity of short-run economic fluctuations.

Stagflation:
A situation of falling output and rising prices; combination of stagnation and inflation.

Steady state:
A condition in which key variables are not changing.

Sticky prices:
Prices that adjust sluggishly and, therefore, do not always equilibrate supply and demand. (Cf. flexible prices.)

Sticky-price model:
The model of aggregate supply emphasizing the slow adjustment of the prices of goods and services.

Stock:
1. A variable measured as a quantity at a point in time. (Cf. flow.) 2. Shares of ownership in a corporation.

Stock market:
A market in which shares of ownership in corporations are bought and sold.

Store of value:
A way of transferring purchasing power from the present to the future; one of the functions of money. (Cf. medium of exchange, unit of account.)

Structural unemployment:
The unemployment resulting from wage rigidity and job rationing. (Cf. frictional unemployment.)

Sub-prime borrower:
A borrower with lower income and assets and thus higher risk of default.

Supply shocks:
Exogenous events that shift the aggregate supply curve.

Tariff:
A tax on imported goods.

Tax multiplier:
The change in aggregate income resulting from a one-dollar change in taxes.

Tax smoothing:
A fiscal policy that aims to keep tax rates stable over time by running budget deficits when government spending is temporarily high or national income is temporarily low.

Taylor principle:
The proposition that a central bank should respond to an increase in inflation with an even greater increase in the nominal interest rate.

Taylor rule:
A rule for monetary policy according to which the central bank sets the interest rate as a function of inflation and the deviation of output from its natural level.

Theory of liquidity preference:
A simple model of the interest rate, based on the ideas in Keynes's *General Theory*, which says that the interest rate adjusts to equilibrate the supply and demand for real money balances.

Time inconsistency:
The tendency of policymakers to announce policies in advance in order to influence the expectations of private decisionmakers, and then to follow different policies after those expectations have been formed and acted upon.

Time-inconsistent preferences:
The possibility of consumers having objectives that change with the passage of time, so they will not follow through on previously made plans.

Tobin's *q*:
The ratio of the market value of installed capital to its replacement cost.

Total factor productivity:
A measure of the level of technology; the amount of output per unit of input, where different inputs are combined on the basis of their factor shares. (Cf. Solow residual.)

Trade balance:
The receipts from exports minus the payments for imports.

Trade deficit:

An excess of imports over exports.

Trade surplus:
An excess of exports over imports.

Transactions velocity of money:
The ratio of the dollar value of all transactions to the money supply.

Transfer payments:
Payments from the government to individuals that are not in exchange for goods and services, such as Social Security payments. (Cf. government purchases.)

Transitory income:
Income that people do not expect to persist into the future; current income minus normal income. (Cf. permanent income.)

Underground economy:
Economic transactions that are hidden in order to evade taxes or conceal illegal activity.

Unemployment insurance:
A government program under which unemployed workers can collect benefits for a certain period of time after losing their jobs.

Unemployment rate:
The percentage of those in the labor force who do not have jobs.

Unit of account:
The measure in which prices and other accounting records are recorded; one of the functions of money. (Cf. medium of exchange, store of value.)

Utility:
A measure of household satisfaction.

Value added:
The value of a firm’s output minus the value of the intermediate goods the firm purchased.

Velocity of money:
The ratio of nominal expenditure to the money supply; the rate at which money changes hands.

Wage:
The amount paid for one unit of labor.

Wage rigidity:
The failure of wages to adjust to equilibrate labor supply and labor demand.

World interest rate:
The interest rate prevailing in world financial markets.

INDEX

Note: Page numbers followed by *f* indicate figures; those followed by *t* indicate tables.

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

Y

Z