openstax™

Principles of

Micro-economics

2e.

for AP® Courses

Principles of Microeconomics for AP® Courses 2e

SENIOR CONTRIBUTING AUTHORS

STEVEN A. GREENLAW, UNIVERSITY OF MARY WASHINGTON

DAVID SHAPIRO, PENNSYLVANIA STATE UNIVERSITY

Based on the 2nd edition of *Principles of Economics, Economics and the Economy, 2e* by Timothy Taylor, published in 2011.

openstax™

OpenStax
Rice University
6100 Main Street MS-375
Houston, Texas 77005

To learn more about OpenStax, visit https://openstax.org.
Individual print copies and bulk orders can be purchased through our website.

Portions of the Demand and Supply chapter were derived from "Why It Matters: Government in Action by Steve Greenlaw and Lumen Learning." Located at: https://courses.lumenlearning.com/waymakermacroxmasterfall2016/chapter/why-it-matters-government-action/. License: CC BY 4.0.

HARDCOVER BOOK ISBN-10	**1-947172-44-1**
HARDCOVER BOOK ISBN-13	**978-1-947172-44-9**
PAPERBACK BOOK ISBN-10	**1-947172-76-X**
PAPERBACK BOOK ISBN-13	**978-1-947172-76-0**
PDF VERSION ISBN-10	**1-947172-45-X**
PDF VERSION ISBN-13	**978-1-947172-45-6**
Revision Number	**MEAC2-2017-000-(11/17)-LC**
Original Publication Year	**2017**

OpenStax

OpenStax provides free, peer-reviewed, openly licensed textbooks for introductory college and Advanced Placement® courses and low-cost, personalized courseware that helps students learn. A nonprofit ed tech initiative based at Rice University, we're committed to helping students access the tools they need to complete their courses and meet their educational goals.

Rice University

OpenStax, OpenStax CNX, and OpenStax Tutor are initiatives of Rice University. As a leading research university with a distinctive commitment to undergraduate education, Rice University aspires to path-breaking research, unsurpassed teaching, and contributions to the betterment of our world. It seeks to fulfill this mission by cultivating a diverse community of learning and discovery that produces leaders across the spectrum of human endeavor.

RICE

Foundation Support

OpenStax is grateful for the tremendous support of our sponsors. Without their strong engagement, the goal of free access to high-quality textbooks would remain just a dream.

Laura and John Arnold Foundation (LJAF) actively seeks opportunities to invest in organizations and thought leaders that have a sincere interest in implementing fundamental changes that not only yield immediate gains, but also repair broken systems for future generations. LJAF currently focuses its strategic investments on education, criminal justice, research integrity, and public accountability.

The William and Flora Hewlett Foundation has been making grants since 1967 to help solve social and environmental problems at home and around the world. The Foundation concentrates its resources on activities in education, the environment, global development and population, performing arts, and philanthropy, and makes grants to support disadvantaged communities in the San Francisco Bay Area.

Calvin K. Kazanjian was the founder and president of Peter Paul (Almond Joy), Inc. He firmly believed that the more people understood about basic economics the happier and more prosperous they would be. Accordingly, he established the Calvin K. Kazanjian Economics Foundation Inc, in 1949 as a philanthropic, nonpolitical educational organization to support efforts that enhanced economic understanding.

Guided by the belief that every life has equal value, the Bill & Melinda Gates Foundation works to help all people lead healthy, productive lives. In developing countries, it focuses on improving people's health with vaccines and other life-saving tools and giving them the chance to lift themselves out of hunger and extreme poverty. In the United States, it seeks to significantly improve education so that all young people have the opportunity to reach their full potential. Based in Seattle, Washington, the foundation is led by CEO Jeff Raikes and Co-chair William H. Gates Sr., under the direction of Bill and Melinda Gates and Warren Buffett.

The Maxfield Foundation supports projects with potential for high impact in science, education, sustainability, and other areas of social importance.

Our mission at The Michelson 20MM Foundation is to grow access and success by eliminating unnecessary hurdles to affordability. We support the creation, sharing, and proliferation of more effective, more affordable educational content by leveraging disruptive technologies, open educational resources, and new models for collaboration between for-profit, nonprofit, and public entities.

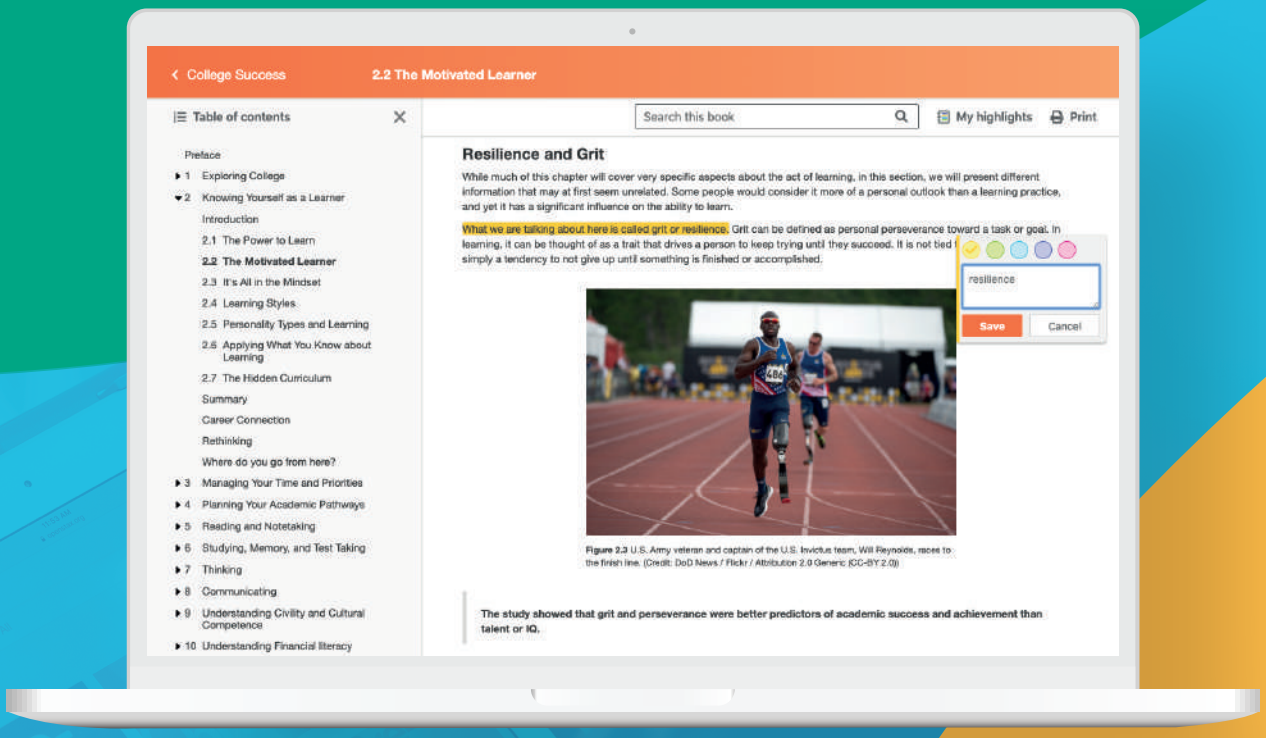
The Bill and Stephanie Sick Fund supports innovative projects in the areas of Education, Art, Science and Engineering.

openstax™

Study where you want, what you want, when you want.

When you access College Success in our web view, you can use our new online **highlighting and note-taking** features to create your own study guides.

Our books are free and flexible, forever.
Get started at **openstax.org/details/books/principles-microeconomics-ap-courses-2e**



Access. The future of education.
openstax.org

RICE

Table of Contents

PREFACE

Welcome to *Principles of Microeconomics for AP® Courses 2e* (2nd Edition), an OpenStax resource. This textbook was written to increase student access to high-quality learning materials, maintaining highest standards of academic rigor at little to no cost.

About OpenStax

OpenStax is a nonprofit based at Rice University, and it's our mission to improve student access to education. Our first openly licensed college textbook was published in 2012, and our library has since scaled to over 25 books for college and AP® courses used by hundreds of thousands of students. OpenStax Tutor, our low-cost personalized learning tool, is being used in college courses throughout the country. Through our partnerships with philanthropic foundations and our alliance with other educational resource organizations, OpenStax is breaking down the most common barriers to learning and empowering students and instructors to succeed.

About OpenStax resources

Customization

Principles of Microeconomics for AP® Courses 2e is licensed under a Creative Commons Attribution 4.0 International (CC BY) license, which means that you can distribute, remix, and build upon the content, as long as you provide attribution to OpenStax and its content contributors.

Because our books are openly licensed, you are free to use the entire book or pick and choose the sections that are most relevant to the needs of your course. Feel free to remix the content by assigning your students certain chapters and sections in your syllabus, in the order that you prefer. You can even provide a direct link in your syllabus to the sections in the web view of your book.

Instructors also have the option of creating a customized version of their OpenStax book. The custom version can be made available to students in low-cost print or digital form through their campus bookstore. Visit the Instructor Resources section of your book page on openstax.org for more information.

Errata

All OpenStax textbooks undergo a rigorous review process. However, like any professional-grade textbook, errors sometimes occur. Since our books are web based, we can make updates periodically when deemed pedagogically necessary. If you have a correction to suggest, submit it through the link on your book page on openstax.org. Subject matter experts review all errata suggestions. OpenStax is committed to remaining transparent about all updates, so you will also find a list of past errata changes on your book page on openstax.org.

Format

You can access this textbook for free in web view or PDF through openstax.org, and for a low cost in print.

About *Principles of Microeconomics for AP® Courses 2e*

Principles of Microeconomics for AP® Courses 2e (2nd edition) covers the scope and sequence of most introductory microeconomics courses. The text includes many current examples, which are handled in a politically equitable way. The outcome is a balanced approach to the theory and application of economics concepts. The second edition has been thoroughly revised to increase clarity, update data and current event impacts, and incorporate the feedback from many reviewers and adopters.

Coverage and scope

To develop the first edition of *Principles of Microeconomics*, we acquired the rights to Timothy Taylor's *Principles of Economics* and solicited ideas from economics instructors at all levels of higher education, from community colleges to PhD-granting universities. For the second edition, we received even more expansive and actionable feedback from hundreds of adopters who had used the book for several academic terms. These knowledgeable instructors informed the pedagogical courses, learning objective development and fulfillment, and the chapter arrangements. Faculty who

taught from the material provided critical and detailed commentary.

The result is a book that covers the breadth of economics topics and also provides the necessary depth to ensure the course is manageable for instructors and students alike. We strove to balance theory and application, as well as the amount of calculation and mathematical examples.

The book is organized into five main parts:

What is Economics? The first two chapters introduce students to the study of economics with a focus on making choices in a world of scarce resources.

Supply and Demand, Chapters 3 and 4, introduces and explains the first analytical model in economics–supply, demand, and equilibrium–before showing applications in the markets for labor and finance.

The Fundamentals of Microeconomic Theory, Chapters 5 through 10, begins the microeconomics portion of the text, presenting the theories of consumer behavior, production and costs, and the different models of market structure, including some simple game theory.

Microeconomic Policy Issues, Chapters 11 through 18, covers the range of topics in applied micro, framed around the concepts of public goods and positive and negative externalities. Students explore competition and antitrust policies, environmental problems, poverty, income inequality, and other labor market issues. The text also covers information, risk and financial markets, and public economy.

International Economics, Chapters 19 and 20, introduces the international dimensions of economics, including international trade and protectionism.

Alternate sequencing

Principles of Microeconomics for AP® Courses 2e was conceived and written to fit a particular topical sequence, but it can be used flexibly to accommodate other course structures. One such potential structure, which fits reasonably well with the textbook content, is provided. Please consider, however, that the chapters were not written to be completely independent, and that the proposed alternate sequence should be carefully considered for student preparation and textual consistency.

Chapter 1 Welcome to Economics!
Chapter 2 Choice in a World of Scarcity
Chapter 3 Demand and Supply
Chapter 4 Labor and Financial Markets
Chapter 5 Elasticity
Chapter 6 Consumer Choices
Chapter 19 International Trade
Chapter 7 Cost and Industry Structure
Chapter 12 Environmental Protection and Negative Externalities
Chapter 13 Positive Externalities and Public Goods
Chapter 8 Perfect Competition
Chapter 9 Monopoly
Chapter 10 Monopolistic Competition and Oligopoly
Chapter 11 Monopoly and Antitrust Policy
Chapter 14 Labor Markets and Income
Chapter 15 Poverty and Economic Inequality
Chapter 16 Information, Risk, and Insurance
Chapter 17 Financial Markets
Chapter 18 Public Economy
Chapter 20 Globalization and Protectionism

Appendix A The Use of Mathematics in Principles of Economics
Appendix B Indifference Curves
Appendix C Present Discounted Value

Changes to the second edition

OpenStax only undertakes revisions when significant modifications to a text are necessary. In the case of *Principles*

of Microeconomics for AP® Courses , we received a wealth of constructive feedback. Many of the book's users felt that consequential movement in economic data, coupled with the impacts of national and global events, warranted a full revision. We also took advantage of the opportunity to improve the writing and sequencing of the text, as well as many of the calculation examples. The major changes are summarized below.

Augmented explanations in chapters one through four provide a more comprehensive and informative foundation for the book.

A clearer explanation, using a numerical example, has been given for finding the utility maximizing combination of goods and services a consumer should choose.

The labor markets chapter and the poverty and economic inequality chapter have been resequenced.

Case studies and examples have been revised and, in some cases, replaced to provide more relevant and useful information for students.

Economic data, tables, and graphs, as well as discussion and analysis around that data, have been thoroughly updated.

Wherever possible, data from the Federal Reserve Economic Database (FRED) was included and referenced. In most of these uses, links to the direct source of the FRED data are provided, and students are encouraged to explore the information and the overall FRED resources more thoroughly.

Additional updates and revisions appear throughout the book. They reflect changes to economic realities and policies regarding international trade, taxation, insurance, and other topics. For issues that may change in the months or years following the textbook's publication, the authors often provided a more open-ended explanation, but we will update the text annually to address further changes.

The revision of *Principles of Microeconomics* was undertaken by Steven Greenlaw (University of Mary Washington) and David Shapiro (Pennsylvania State University), with significant input by lead reviewer Daniel MacDonald (California State University, San Bernardino).

Pedagogical foundation

Throughout *Principles of Microeconomics for AP® Courses 2e*, you will find features that engage the students in economic inquiry and support the understanding required in the AP® course. Our features include:

Bring It Home. This feature presents a brief case study, specific to each chapter, which connects the chapter's main topic to the real world. It is broken up into two parts: the first at the beginning of the chapter (in the Intro module) and the second at chapter's end, when students have learned what's necessary to understand the case and "bring home" the chapter's core concepts.

Work It Out. This feature asks students to work through a generally analytical or computational problem, and guides them step-by-step to find out how its solution is derived.

Clear It Up. This feature addresses common student misconceptions about the content. Clear It Ups are usually deeper explanations of something in the main body of the text. Each CIU starts with a question. The rest of the feature explains the answer.

Link It Up. This feature offers a very brief introduction to a website that is pertinent to students' understanding and enjoyment of the topic at hand.

Questions for each level of learning

Principles of Microeconomics for AP® Courses 2e offers four types of end-of-module questions for students.

Self-Checks. These analytical self-assessment questions appear at the end of each module. They push the student to think beyond what is said in the text. Self-Check questions are designed for formative (rather than summative) assessment. The questions and answers are explained so that students feel like they are being walked through the problem.

Review Questions. These are simple recall questions from the chapter and are in open-response format (not multiple choice or true/false). The answers can be looked up in the text.

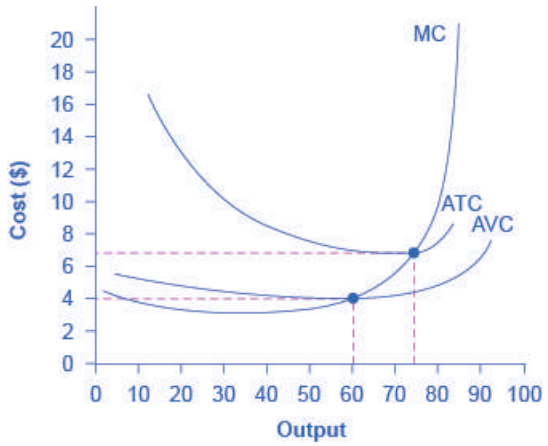
Critical Thinking Questions. These higher-level, conceptual questions ask students to *demonstrate their understanding by applying* what they have learned in different contexts. They ask for *reasoning* about the

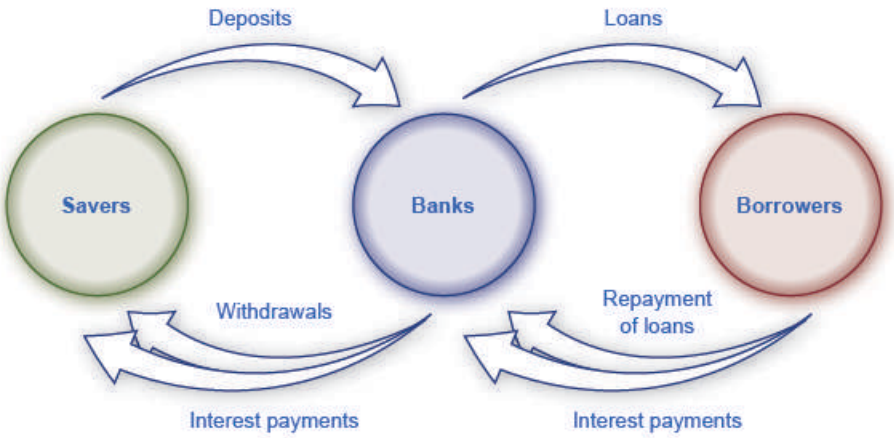
concepts in a manner that will help prepare students for the AP® exam.

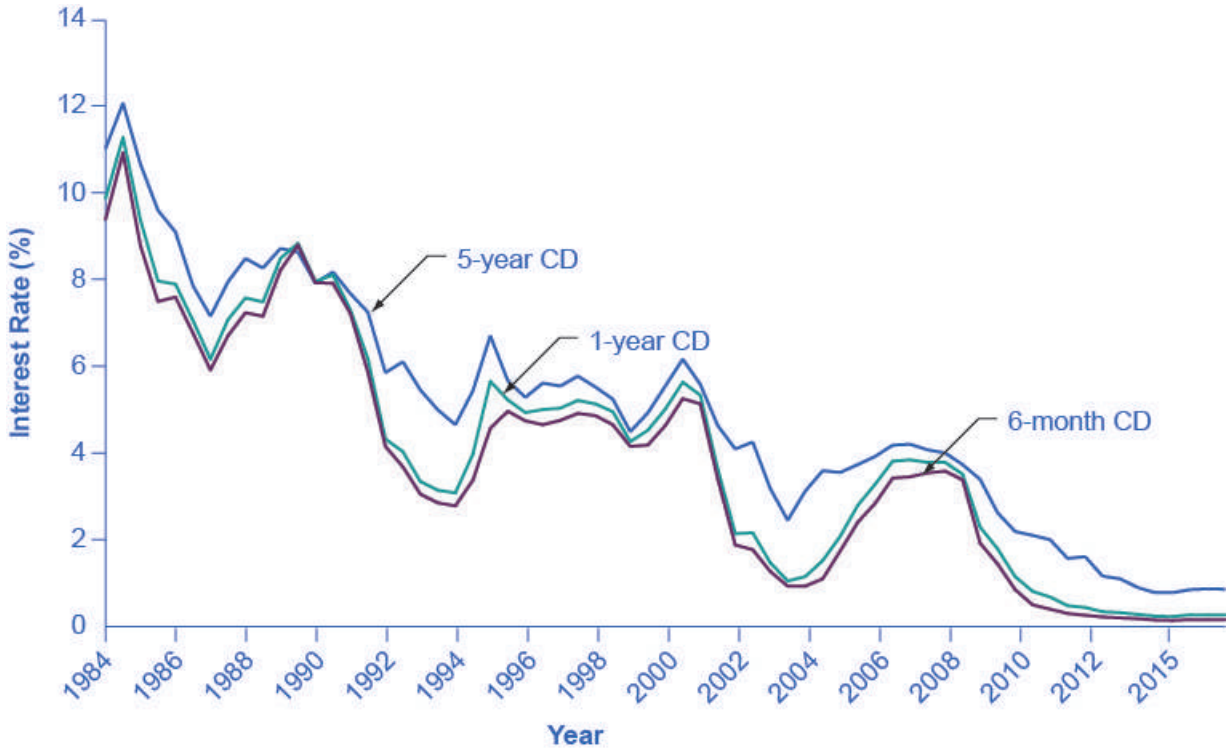
Problems. These exercises give students additional practice working with the analytic and computational concepts in the module.

Updated art

Principles of Microeconomics for AP® Courses 2e includes an updated art program to better inform today's student, providing the latest data on covered topics.







Additional resources

Student and instructor resources

We've compiled additional resources for both students and instructors, including Getting Started Guides, an instructor solution manual, test bank, and PowerPoint slides. Instructor resources require a verified instructor account, which you can apply for when you log in or create your account on openstax.org. Take advantage of these resources to supplement your OpenStax book.

Community Hubs

OpenStax partners with the Institute for the Study of Knowledge Management in Education (ISKME) to offer Community Hubs on OER Commons – a platform for instructors to share community-created resources that support OpenStax books, free of charge. Through our Community Hubs, instructors can upload their own materials or download resources to use in their own courses, including additional ancillaries, teaching material, multimedia, and relevant course content. We encourage instructors to join the hubs for the subjects most relevant to your teaching and research as an opportunity both to enrich your courses and to engage with other faculty.

To reach the Community Hubs, visit www.oercommons.org/hubs/OpenStax.

Technology partners

As allies in making high-quality learning materials accessible, our technology partners offer optional low-cost tools that are integrated with OpenStax books. To access the technology options for your text, visit your book page on openstax.org.

About the authors

Senior contributing authors

Steven A. Greenlaw, University of Mary Washington
Steven Greenlaw has been teaching principles of economics for more than 30 years. In 1999, he received the Grellet C. Simpson Award for Excellence in Undergraduate Teaching at the University of Mary Washington. He is the author of *Doing Economics: A Guide to Doing and Understanding Economic Research*, as well as a variety of articles on economics pedagogy and instructional technology, published in the *Journal of Economic Education*, the *International Review of Economic Education*, and other outlets. He wrote the module on Quantitative Writing for *Starting Point: Teaching and Learning Economics*, the web portal on best practices in teaching economics. Steven Greenlaw lives in Alexandria, Virginia with his wife Kathy and their three children.

David Shapiro, Pennsylvania State University
David Shapiro is Professor Emeritus of Economics, Demography, and Women's, Gender, and Sexuality Studies at the Pennsylvania State University. He received a BA in Economics and Political Science from the University of Michigan, and an MA as well as a PhD in Economics from Princeton University. He began his academic career at Ohio State University in 1971, and moved to Penn State in 1980. His early research focused on women and youth in the United States labor market. Following a 1978-79 stint as a Fulbright professor at the University of Kinshasa in the Democratic Republic of the Congo, his research shifted focus to fertility in Kinshasa and more broadly, in sub-Saharan Africa. He has also received the top prize for teaching at both Ohio State and Penn State.

Special thanks to Christian Potter from University of Mary Washington, who thoroughly researched and applied many of the data updates and provided the foundation for many new and revised illustrations.

Development editor

Thomas Sigel

Contributing authors

Eric Dodge, Hanover College
Cynthia Gamez, University of Texas at El Paso
Andres Jauregui, Columbus State University
Diane Keenan, Cerritos College
Dan MacDonald, California State University, San Bernardino
Amyaz Moledina, The College of Wooster
Craig Richardson, Winston-Salem State University
Ralph Sonenshine, American University

Reviewers

Bryan Aguiar, Northwest Arkansas Community College
Basil Al Hashimi, Mesa Community College
Jennifer Ball, Washburn University
Emil Berendt, Mount St. Mary's University
Zena Buser, Adams State University
Douglas Campbell, The University of Memphis
Sanjukta Chaudhuri, University of Wisconsin-Eau Claire
Xueyu Cheng, Alabama State University

Robert Cunningham, Alma College
Rosa Lea Danielson, College of DuPage
Steven Deloach, Elon University
Debbie Evercloud, University of Colorado Denver
Sal Figueras, Hudson County Community College
Reza Ghorashi, Richard Stockton College of New Jersey
Robert Gillette, University of Kentucky
George Jones, University of Wisconsin-Rock County
Charles Kroncke, College of Mount St. Joseph
Teresa Laughlin, Palomar Community College
Carlos Liard-Muriente, Central Connecticut State University
Heather Luea, Kansas State University
Charles Meyrick, Housatonic Community College
William Mosher, Nashua Community College
Michael Netta, Hudson County Community College
Nick Noble, Miami University
Joe Nowakowski, Muskingum University
Shawn Osell, University of Wisconsin, Superior
Mark Owens, Middle Tennessee State University
Sonia Pereira, Barnard College
Brian Peterson, Central College
Jennifer Platania, Elon University
Robert Rycroft, University of Mary Washington
Adrienne Sachse, Florida State College at Jacksonville
Hans Schumann, Texas A&M University
Gina Shamshak, Goucher College
Chris Warburton, John Jay College of Criminal Justice, CUNY
Mark Witte, Northwestern University
Chiou-nan Yeh, Alabama State University

1 | Welcome to Economics!

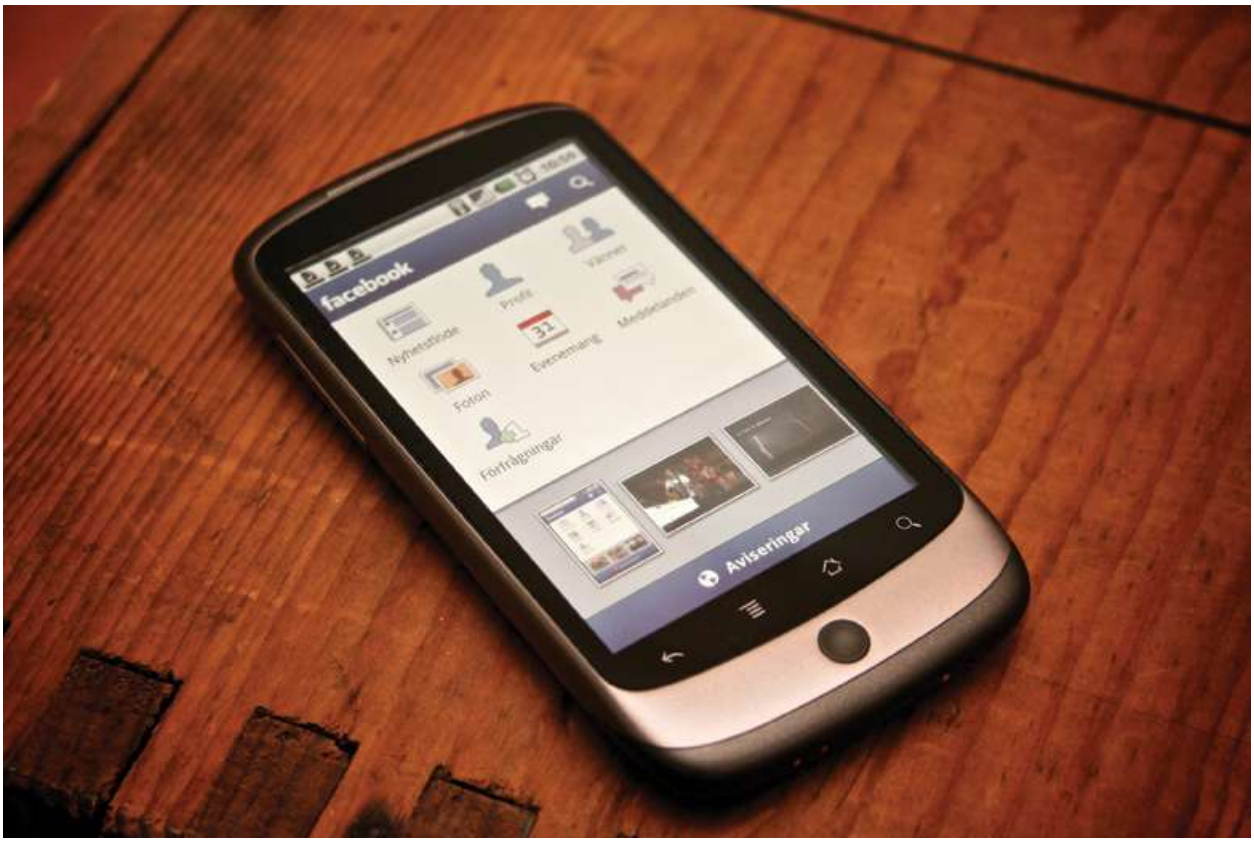


Figure 1.1 Do You Use Facebook? Economics is greatly impacted by how well information travels through society. Today, social media giants Twitter, Facebook, and Instagram are major forces on the information super highway. (Credit: Johan Larsson/Flickr)

Bring it Home

Decisions ... Decisions in the Social Media Age

To post or not to post? Every day we are faced with a myriad of decisions, from what to have for breakfast, to which route to take to class, to the more complex—"Should I double major and add possibly another semester of study to my education?" Our response to these choices depends on the information we have available at any given moment. Economists call this "imperfect" because we rarely have all the data we need to make perfect decisions. Despite the lack of perfect information, we still make hundreds of decisions a day.

Now we have another avenue in which to gather information—social media. Outlets like Facebook and Twitter are altering the process by which we make choices, how we spend our time, which movies we see, which products we buy, and more. How many of you chose a university without checking out its Facebook page or Twitter stream first for information and feedback?

As you will see in this course, what happens in economics is affected by how well and how fast information disseminates through a society, such as how quickly information travels through Facebook. "Economists love nothing better than when deep and liquid markets operate under conditions of perfect information," says Jessica Irvine, National Economics Editor for News Corp Australia.

This leads us to the topic of this chapter, an introduction to the world of making decisions, processing

information, and understanding behavior in markets —the world of economics. Each chapter in this book will start with a discussion about current (or sometimes past) events and revisit it at chapter's end—to "bring home" the concepts in play.

Introduction

In this chapter, you will learn about:

- What Is Economics, and Why Is It Important?
- Microeconomics and Macroeconomics
- How Economists Use Theories and Models to Understand Economic Issues
- How Economies Can Be Organized: An Overview of Economic Systems

What is economics and why should you spend your time learning it? After all, there are other disciplines you could be studying, and other ways you could be spending your time. As the Bring it Home feature just mentioned, making choices is at the heart of what economists study, and your decision to take this course is as much as economic decision as anything else.

Economics is probably not what you think. It is not primarily about money or finance. It is not primarily about business. It is not mathematics. What is it then? It is both a subject area and a way of viewing the world.

1.1 | What Is Economics, and Why Is It Important?

By the end of this section, you will be able to:

- Discuss the importance of studying economics
- Explain the relationship between production and division of labor
- Evaluate the significance of scarcity

Economics is the study of how humans make decisions in the face of scarcity. These can be individual decisions, family decisions, business decisions or societal decisions. If you look around carefully, you will see that scarcity is a fact of life. **Scarcity** means that human wants for goods, services and resources exceed what is available. Resources, such as labor, tools, land, and raw materials are necessary to produce the goods and services we want but they exist in limited supply. Of course, the ultimate scarce resource is time- everyone, rich or poor, has just 24 expendable hours in the day to earn income to acquire goods and services, for leisure time, or for sleep. At any point in time, there is only a finite amount of resources available.

Think about it this way: In 2015 the labor force in the United States contained over 158 million workers, according to the U.S. Bureau of Labor Statistics. The total land area was 3,794,101 square miles. While these are certainly large numbers, they are not infinite. Because these resources are limited, so are the numbers of goods and services we produce with them. Combine this with the fact that human wants seem to be virtually infinite, and you can see why scarcity is a problem.

Introduction to FRED

Data is very important in economics because it describes and measures the issues and problems that economics seek to understand. A variety of government agencies publish economic and social data. For this course, we will generally use data from the St. Louis Federal Reserve Bank's FRED database. FRED is very user friendly. It allows you to display data in tables or charges, and you can easily download it into spreadsheet form if you want to use the data for other purposes. The **FRED website (https://openstax.org/l/FRED/)** includes data on nearly 400,000 domestic and international variables over time, in the following broad categories:

- Money, Banking & Finance
- Population, Employment, & Labor Markets (including Income Distribution)

- National Accounts (Gross Domestic Product & its components), Flow of Funds, and International Accounts
- Production & Business Activity (including Business Cycles)
- Prices & Inflation (including the Consumer Price Index, the Producer Price Index, and the Employment Cost Index)
- International Data from other nations
- U.S. Regional Data
- Academic Data (including Penn World Tables & NBER Macrohistory database)

For more information about how to use FRED, see the variety of **videos (https://openstax.org/l/FRED_intro)** on YouTube starting with this introduction.

Figure 1.2 Scarcity of Resources Homeless people are a stark reminder that scarcity of resources is real. (Credit: "daveynin"/Flickr Creative Commons)

If you still do not believe that scarcity is a problem, consider the following: Does everyone require food to eat? Does everyone need a decent place to live? Does everyone have access to healthcare? In every country in the world, there are people who are hungry, homeless (for example, those who call park benches their beds, as **Figure 1.2** shows), and in need of healthcare, just to focus on a few critical goods and services. Why is this the case? It is because of scarcity. Let's delve into the concept of scarcity a little deeper, because it is crucial to understanding economics.

The Problem of Scarcity

Think about all the things you consume: food, shelter, clothing, transportation, healthcare, and entertainment. How do you acquire those items? You do not produce them yourself. You buy them. How do you afford the things you buy? You work for pay. If you do not, someone else does on your behalf. Yet most of us never have enough income to buy all the things we want. This is because of scarcity. So how do we solve it?

Link It Up

Visit this website (http://openstax.org/l/drought) to read about how the United States is dealing with scarcity in resources.

Every society, at every level, must make choices about how to use its resources. Families must decide whether to spend their money on a new car or a fancy vacation. Towns must choose whether to put more of the budget into police and fire protection or into the school system. Nations must decide whether to devote more funds to national defense or to protecting the environment. In most cases, there just isn't enough money in the budget to do everything. How do we use our limited resources the best way possible, that is, to obtain the most goods and services we can? There are a couple of options. First, we could each produce everything we each consume. Alternatively, we could each produce some of what we want to consume, and "trade" for the rest of what we want. Let's explore these options. Why do we not each just produce all of the things we consume? Think back to pioneer days, when individuals knew how to do so much more than we do today, from building their homes, to growing their crops, to hunting for food, to repairing their equipment. Most of us do not know how to do all—or any—of those things, but it is not because we could not learn. Rather, we do not have to. The reason why is something called *the division and specialization of labor*, a production innovation first put forth by Adam Smith (**Figure 1.3**) in his book, *The Wealth of Nations*.

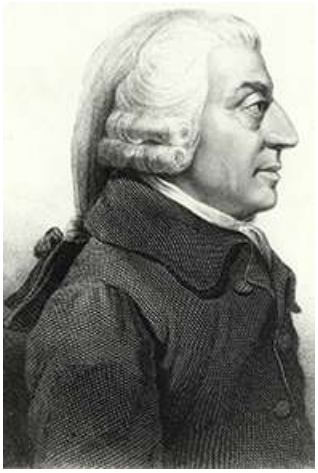


Figure 1.3 Adam Smith Adam Smith introduced the idea of dividing labor into discrete tasks. (Credit: Wikimedia Commons)

The Division of and Specialization of Labor

The formal study of economics began when Adam Smith (1723–1790) published his famous book *The Wealth of Nations* in 1776. Many authors had written on economics in the centuries before Smith, but he was the first to address the subject in a comprehensive way. In the first chapter, Smith introduces the concept of **division of labor**, which means that the way one produces a good or service is divided into a number of tasks that different workers perform, instead of all the tasks being done by the same person.

To illustrate division of labor, Smith counted how many tasks went into making a pin: drawing out a piece of wire, cutting it to the right length, straightening it, putting a head on one end and a point on the other, and packaging pins for sale, to name just a few. Smith counted 18 distinct tasks that different people performed—all for a pin, believe it or not!

Modern businesses divide tasks as well. Even a relatively simple business like a restaurant divides the task of serving meals into a range of jobs like top chef, sous chefs, less-skilled kitchen help, servers to wait on the tables, a greeter at the door, janitors to clean up, and a business manager to handle paychecks and bills—not to mention the economic

connections a restaurant has with suppliers of food, furniture, kitchen equipment, and the building where it is located. A complex business like a large manufacturing factory, such as the shoe factory (**Figure 1.4**), or a hospital can have hundreds of job classifications.

Figure 1.4 Division of Labor Workers on an assembly line are an example of the divisions of labor. (Credit: Nina Hale/Flickr Creative Commons)

Why the Division of Labor Increases Production

When we divide and subdivide the tasks involved with producing a good or service, workers and businesses can produce a greater quantity of output. In his observations of pin factories, Smith noticed that one worker alone might make 20 pins in a day, but that a small business of 10 workers (some of whom would need to complete two or three of the 18 tasks involved with pin-making), could make 48,000 pins in a day. How can a group of workers, each specializing in certain tasks, produce so much more than the same number of workers who try to produce the entire good or service by themselves? Smith offered three reasons.

First, **specialization** in a particular small job allows workers to focus on the parts of the production process where they have an advantage. (In later chapters, we will develop this idea by discussing comparative advantage.) People have different skills, talents, and interests, so they will be better at some jobs than at others. The particular advantages may be based on educational choices, which are in turn shaped by interests and talents. Only those with medical degrees qualify to become doctors, for instance. For some goods, geography affects specialization. For example, it is easier to be a wheat farmer in North Dakota than in Florida, but easier to run a tourist hotel in Florida than in North Dakota. If you live in or near a big city, it is easier to attract enough customers to operate a successful dry cleaning business or movie theater than if you live in a sparsely populated rural area. Whatever the reason, if people specialize in the production of what they do best, they will be more effective than if they produce a combination of things, some of which they are good at and some of which they are not.

Second, workers who specialize in certain tasks often learn to produce more quickly and with higher quality. This pattern holds true for many workers, including assembly line laborers who build cars, stylists who cut hair, and doctors who perform heart surgery. In fact, specialized workers often know their jobs well enough to suggest innovative ways to do their work faster and better.

A similar pattern often operates within businesses. In many cases, a business that focuses on one or a few products (sometimes called its “core competency”) is more successful than firms that try to make a wide range of products.

Third, specialization allows businesses to take advantage of **economies of scale**, which means that for many goods, as the level of production increases, the average cost of producing each individual unit declines. For example, if a factory produces only 100 cars per year, each car will be quite expensive to make on average. However, if a factory produces 50,000 cars each year, then it can set up an assembly line with huge machines and workers performing specialized tasks, and the average cost of production per car will be lower. The ultimate result of workers who can focus on their preferences and talents, learn to do their specialized jobs better, and work in larger organizations is that society as a whole can produce and consume far more than if each person tried to produce all of his or her own goods and services. The division and specialization of labor has been a force against the problem of scarcity.

Trade and Markets

Specialization only makes sense, though, if workers can use the pay they receive for doing their jobs to purchase the other goods and services that they need. In short, specialization requires trade.

You do not have to know anything about electronics or sound systems to play music—you just buy an iPod or MP3 player, download the music, and listen. You do not have to know anything about artificial fibers or the construction of sewing machines if you need a jacket—you just buy the jacket and wear it. You do not need to know anything about internal combustion engines to operate a car—you just get in and drive. Instead of trying to acquire all the knowledge and skills involved in producing all of the goods and services that you wish to consume, the market allows you to learn a specialized set of skills and then use the pay you receive to buy the goods and services you need or want. This is how our modern society has evolved into a strong economy.

Why Study Economics?

Now that you have an overview on what economics studies, let's quickly discuss why you are right to study it. Economics is not primarily a collection of facts to memorize, although there are plenty of important concepts to learn. Instead, think of economics as a collection of questions to answer or puzzles to work. Most importantly, economics provides the tools to solve those puzzles. If the economics "bug" has not bitten you yet, there are other reasons why you should study economics.

- Virtually every major problem facing the world today, from global warming, to world poverty, to the conflicts in Syria, Afghanistan, and Somalia, has an economic dimension. If you are going to be part of solving those problems, you need to be able to understand them. Economics is crucial.
- It is hard to overstate the importance of economics to good citizenship. You need to be able to vote intelligently on budgets, regulations, and laws in general. When the U.S. government came close to a standstill at the end of 2012 due to the "fiscal cliff," what were the issues? Did you know?
- A basic understanding of economics makes you a well-rounded thinker. When you read articles about economic issues, you will understand and be able to evaluate the writer's argument. When you hear classmates, co-workers, or political candidates talking about economics, you will be able to distinguish between common sense and nonsense. You will find new ways of thinking about current events and about personal and business decisions, as well as current events and politics.

The study of economics does not dictate the answers, but it can illuminate the different choices.

1.2 | Microeconomics and Macroeconomics

By the end of this section, you will be able to:

- Describe microeconomics
- Describe macroeconomics
- Contrast monetary policy and fiscal policy

Economics is concerned with the well-being of *all* people, including those with jobs and those without jobs, as well as those with high incomes and those with low incomes. Economics acknowledges that production of useful goods and services can create problems of environmental pollution. It explores the question of how investing in education helps to develop workers' skills. It probes questions like how to tell when big businesses or big labor unions are operating in a way that benefits society as a whole and when they are operating in a way that benefits their owners or members at the expense of others. It looks at how government spending, taxes, and regulations affect decisions about production and consumption.

It should be clear by now that economics covers considerable ground. We can divide that ground into two parts: **Microeconomics** focuses on the actions of individual agents within the economy, like households, workers, and businesses. **Macroeconomics** looks at the economy as a whole. It focuses on broad issues such as growth of production, the number of unemployed people, the inflationary increase in prices, government deficits, and levels of exports and imports. Microeconomics and macroeconomics are not separate subjects, but rather complementary perspectives on the overall subject of the economy.

To understand why both microeconomic and macroeconomic perspectives are useful, consider the problem of studying a biological ecosystem like a lake. One person who sets out to study the lake might focus on specific topics: certain kinds of algae or plant life; the characteristics of particular fish or snails; or the trees surrounding the lake. Another person might take an overall view and instead consider the lake's ecosystem from top to bottom; what eats what, how the system stays in a rough balance, and what environmental stresses affect this balance. Both approaches are useful, and both examine the same lake, but the viewpoints are different. In a similar way, both microeconomics and macroeconomics study the same economy, but each has a different viewpoint.

Whether you are scrutinizing lakes or economics, the micro and the macro insights should blend with each other. In studying a lake, the micro insights about particular plants and animals help to understand the overall food chain, while the macro insights about the overall food chain help to explain the environment in which individual plants and animals live.

In economics, the micro decisions of individual businesses are influenced by whether the macroeconomy is healthy. For example, firms will be more likely to hire workers if the overall economy is growing. In turn, macroeconomy's performance ultimately depends on the microeconomic decisions that individual households and businesses make.

Microeconomics

What determines how households and individuals spend their budgets? What combination of goods and services will best fit their needs and wants, given the budget they have to spend? How do people decide whether to work, and if so, whether to work full time or part time? How do people decide how much to save for the future, or whether they should borrow to spend beyond their current means?

What determines the products, and how many of each, a firm will produce and sell? What determines the prices a firm will charge? What determines how a firm will produce its products? What determines how many workers it will hire? How will a firm finance its business? When will a firm decide to expand, downsize, or even close? In the microeconomics part of this book, we will learn about the theory of consumer behavior, the theory of the firm, how markets for labor and other resources work, and how markets sometimes fail to work properly.

Macroeconomics

What determines the level of economic activity in a society? In other words, what determines how many goods and services a nation actually produces? What determines how many jobs are available in an economy? What determines a nation’s standard of living? What causes the economy to speed up or slow down? What causes firms to hire more workers or to lay them off? Finally, what causes the economy to grow over the long term?

We can determine an economy's macroeconomic health by examining a number of goals: growth in the standard of living, low unemployment, and low inflation, to name the most important. How can we use government macroeconomic policy to pursue these goals? A nation's central bank conducts **monetary policy**, which involves policies that affect bank lending, interest rates, and financial capital markets. For the United States, this is the Federal Reserve. A nation's legislative body determines **fiscal policy**, which involves government spending and taxes. For the United States, this is the Congress and the executive branch, which originates the federal budget. These are the government's main tools. Americans tend to expect that government can fix whatever economic problems we encounter, but to what extent is that expectation realistic? These are just some of the issues that we will explore in the macroeconomic chapters of this book.

1.3 | How Economists Use Theories and Models to Understand Economic Issues

By the end of this section, you will be able to:

- Interpret a circular flow diagram
- Explain the importance of economic theories and models
- Describe goods and services markets and labor markets

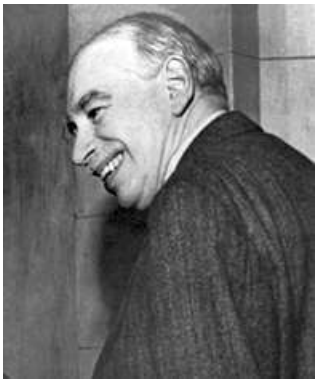


Figure 1.5 John Maynard Keynes One of the most influential economists in modern times was John Maynard Keynes. (Credit: Wikimedia Commons)

John Maynard Keynes (1883–1946), one of the greatest economists of the twentieth century, pointed out that economics is not just a subject area but also a way of thinking. Keynes (**Figure 1.5**) famously wrote in the introduction to a fellow economist's book: "[Economics] is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions." In other words, economics teaches you how to think, not what to think.

Link It Up

Watch this video (http://openstax.org/l/Keynes) about John Maynard Keynes and his influence on economics.

Economists see the world through a different lens than anthropologists, biologists, classicists, or practitioners of any other discipline. They analyze issues and problems using economic theories that are based on particular assumptions about human behavior. These assumptions tend to be different than the assumptions an anthropologist or psychologist might use. A **theory** is a simplified representation of how two or more variables interact with each other. The purpose of a theory is to take a complex, real-world issue and simplify it down to its essentials. If done well, this enables the analyst to understand the issue and any problems around it. A good theory is simple enough to understand, while complex enough to capture the key features of the object or situation you are studying.

Sometimes economists use the term **model** instead of theory. Strictly speaking, a theory is a more abstract representation, while a model is a more applied or empirical representation. We use models to test theories, but for this course we will use the terms interchangeably.

For example, an architect who is planning a major office building will often build a physical model that sits on a tabletop to show how the entire city block will look after the new building is constructed. Companies often build models of their new products, which are more rough and unfinished than the final product, but can still demonstrate how the new product will work.

A good model to start with in economics is the **circular flow diagram** (**Figure 1.6**). It pictures the economy as consisting of two groups—households and firms—that interact in two markets: the **goods and services market** in which firms sell and households buy and the **labor market** in which households sell labor to business firms or other employees.

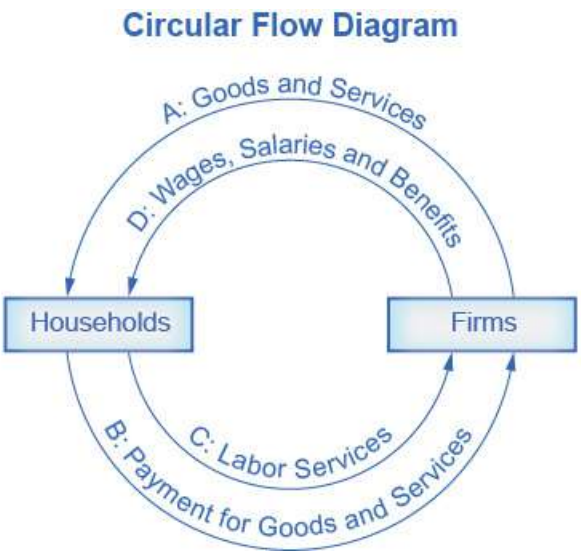


Figure 1.6 The Circular Flow Diagram The circular flow diagram shows how households and firms interact in the goods and services market, and in the labor market. The direction of the arrows shows that in the goods and services market, households receive goods and services and pay firms for them. In the labor market, households provide labor and receive payment from firms through wages, salaries, and benefits.

Firms produce and sell goods and services to households in the market for goods and services (or product market). Arrow “A” indicates this. Households pay for goods and services, which becomes the revenues to firms. Arrow “B” indicates this. Arrows A and B represent the two sides of the product market. Where do households obtain the income to buy goods and services? They provide the labor and other resources (e.g. land, capital, raw materials) firms need to produce goods and services in the market for inputs (or factors of production). Arrow “C” indicates this. In return, firms pay for the inputs (or resources) they use in the form of wages and other factor payments. Arrow “D” indicates this. Arrows “C” and “D” represent the two sides of the factor market.

Of course, in the real world, there are many different markets for goods and services and markets for many different types of labor. The circular flow diagram simplifies this to make the picture easier to grasp. In the diagram, firms produce goods and services, which they sell to households in return for revenues. The outer circle shows this, and represents the two sides of the product market (for example, the market for goods and services) in which households demand and firms supply. Households sell their labor as workers to firms in return for wages, salaries, and benefits. The inner circle shows this and represents the two sides of the labor market in which households supply and firms demand.

This version of the circular flow model is stripped down to the essentials, but it has enough features to explain how the product and labor markets work in the economy. We could easily add details to this basic model if we wanted to introduce more real-world elements, like financial markets, governments, and interactions with the rest of the globe (imports and exports).

Economists carry a set of theories in their heads like a carpenter carries around a toolkit. When they see an economic issue or problem, they go through the theories they know to see if they can find one that fits. Then they use the theory to derive insights about the issue or problem. Economists express theories as diagrams, graphs, or even as mathematical equations. (Do not worry. In this course, we will mostly use graphs.) Economists do not figure out the answer to the problem first and then draw the graph to illustrate. Rather, they use the graph of the theory to help them figure out the answer. Although at the introductory level, you can sometimes figure out the right answer without applying a model, if you keep studying economics, before too long you will run into issues and problems that you will need to graph to solve. We explain both micro and macroeconomics in terms of theories and models. The most well-known theories are probably those of supply and demand, but you will learn a number of others.

1.4 | How To Organize Economies: An Overview of Economic Systems

By the end of this section, you will be able to:

- Contrast traditional economies, command economies, and market economies
- Explain gross domestic product (GDP)
- Assess the importance and effects of globalization

Think about what a complex system a modern economy is. It includes all production of goods and services, all buying and selling, all employment. The economic life of every individual is interrelated, at least to a small extent, with the economic lives of thousands or even millions of other individuals. Who organizes and coordinates this system? Who insures that, for example, the number of televisions a society provides is the same as the amount it needs and wants? Who insures that the right number of employees work in the electronics industry? Who insures that televisions are produced in the best way possible? How does it all get done?

There are at least three ways that societies organize an economy. The first is the **traditional economy**, which is the oldest economic system and is used in parts of Asia, Africa, and South America. Traditional economies organize their economic affairs the way they have always done (i.e., tradition). Occupations stay in the family. Most families are farmers who grow the crops using traditional methods. What you produce is what you consume. Because tradition drives the way of life, there is little economic progress or development.

Figure 1.7 A Command Economy Ancient Egypt was an example of a command economy. (Credit: Jay Bergesen/ Flickr Creative Commons)

Command economies are very different. In a **command economy**, economic effort is devoted to goals passed down from a ruler or ruling class. Ancient Egypt was a good example: a large part of economic life was devoted to building pyramids, like those in **Figure 1.7**, for the pharaohs. Medieval manor life is another example: the lord provided the land for growing crops and protection in the event of war. In return, vassals provided labor and soldiers to do the lord's bidding. In the last century, communism emphasized command economies.

In a command economy, the government decides what goods and services will be produced and what prices it will charge for them. The government decides what methods of production to use and sets wages for workers. The government provides many necessities like healthcare and education for free. Currently, Cuba and North Korea have command economies.

Figure 1.8 A Market Economy Nothing says “market” more than The New York Stock Exchange. (Credit: Erik Drost/ Flickr Creative Commons)

Although command economies have a very centralized structure for economic decisions, market economies have a very decentralized structure. A **market** is an institution that brings together buyers and sellers of goods or services, who may be either individuals or businesses. The New York Stock Exchange (**Figure 1.8**) is a prime example of a market which brings buyers and sellers together. In a **market economy**, decision-making is decentralized. Market economies are based on **private enterprise**: the private individuals or groups of private individuals own and operate the means of production (resources and businesses). Businesses supply goods and services based on demand. (In a command economy, by contrast, the government owns resources and businesses.) Supply of goods and services depends on what the demands. A person’s income is based on his or her ability to convert resources (especially labor) into something that society values. The more society values the person’s output, the higher the income (think Lady Gaga or LeBron James). In this scenario, market forces, not governments, determine economic decisions.

Most economies in the real world are mixed. They combine elements of command and market (and even traditional) systems. The U.S. economy is positioned toward the market-oriented end of the spectrum. Many countries in Europe and Latin America, while primarily market-oriented, have a greater degree of government involvement in economic decisions than the U.S. economy. China and Russia, while over the past several decades have moved more in the direction of having a market-oriented system, remain closer to the command economy end of the spectrum. The Heritage Foundation provides information about how free and thus market-oriented different countries' are, as the following Clear It Up feature discusses. For a similar ranking, but one that defines freedom more broadly, see the Cato Foundation's Human Freedom **Index (https://openstax.org/l/cato)** .

Clear It Up

What countries are considered economically free?

Who is in control of economic decisions? Are people free to do what they want and to work where they want? Are businesses free to produce when they want and what they choose, and to hire and fire as they wish? Are banks free to choose who will receive loans, or does the government control these kinds of choices? Each year, researchers at the Heritage Foundation and the *Wall Street Journal* look at 50 different categories of economic freedom for countries around the world. They give each nation a score based on the extent of economic freedom in each category.

The 2016 Heritage Foundation’s Index of Economic Freedom report ranked 178 countries around the world: **Table 1.1** lists some examples of the most free and the least free countries. Several additional countries were not ranked because of extreme instability that made judgments about economic freedom impossible. These countries include Afghanistan, Iraq, Libya, Syria, Somalia, and Yemen.

The assigned rankings are inevitably based on estimates, yet even these rough measures can be useful for discerning trends. In 2015, 101 of the 178 included countries shifted toward greater economic freedom, although 77 of the countries shifted toward less economic freedom. In recent decades, the overall trend has been a *higher level of economic freedom around the world*.

Most Economic Freedom	Least Economic Freedom
1. Hong Kong	167. Timor-Leste
2. Singapore	168. Democratic Republic of Congo
3. New Zealand	169. Argentina
4. Switzerland	170. Equatorial Guinea
5. Australia	171. Iran
6. Canada	172. Republic of Congo
7. Chile	173. Eritrea
8. Ireland	174. Turkmenistan
9. Estonia	175. Zimbabwe
10. United Kingdom	176. Venezuela
11. United States	177. Cuba
12. Denmark	178. North Korea

Table 1.1 Economic Freedoms, 2016 (Source: The Heritage Foundation, 2016 Index of Economic Freedom, Country Rankings, http://www.heritage.org/index/ranking)

Regulations: The Rules of the Game

Markets and government regulations are always entangled. There is no such thing as an absolutely free market. Regulations always define the "rules of the game" in the economy. Economies that are primarily market-oriented have fewer regulations—ideally just enough to maintain an even playing field for participants. At a minimum, these laws govern matters like safeguarding private property against theft, protecting people from violence, enforcing legal contracts, preventing fraud, and collecting taxes. Conversely, even the most command-oriented economies operate using markets. How else would buying and selling occur? The government heavily regulates decisions of what to produce and prices to charge. Heavily regulated economies often have **underground economies** (or black markets), which are markets where the buyers and sellers make transactions without the government's approval.

The question of how to organize economic institutions is typically not a black-or-white choice between all market or all government, but instead involves a balancing act over the appropriate combination of market freedom and government rules.

Figure 1.9 Globalization Cargo ships are one mode of transportation for shipping goods in the global economy. (Credit: Raul Valdez/Flickr Creative Commons)

The Rise of Globalization

Recent decades have seen a trend toward **globalization**, which is the expanding cultural, political, and economic connections between people around the world. One measure of this is the increased buying and selling of goods, services, and assets across national borders—in other words, international trade and financial capital flows.

Globalization has occurred for a number of reasons. Improvements in shipping, as illustrated by the container ship in **Figure 1.9**, and air cargo have driven down transportation costs. Innovations in computing and telecommunications have made it easier and cheaper to manage long-distance economic connections of production and sales. Many valuable products and services in the modern economy can take the form of information—for example: computer software; financial advice; travel planning; music, books and movies; and blueprints for designing a building. These products and many others can be transported over telephones and computer networks at ever-lower costs. Finally, international agreements and treaties between countries have encouraged greater trade.

Table 1.2 presents one measure of globalization. It shows the percentage of domestic economic production that was exported for a selection of countries from 2010 to 2015, according to an entity known as The World Bank. **Exports** are the goods and services that one produces domestically and sells abroad. **Imports** are the goods and services that one produces abroad and then sells domestically. **Gross domestic product (GDP)** measures the size of total production in an economy. Thus, the ratio of exports divided by GDP measures what share of a country's total economic production is sold in other countries.

Country	2010	2011	2012	2013	2014	2015
Higher Income Countries						
United States	12.4	13.6	13.6	13.5	13.5	12.6
Belgium	76.2	81.4	82.2	82.8	84.0	84.4
Canada	29.1	30.7	30.0	30.1	31.7	31.5
France	26.0	27.8	28.1	28.3	29.0	30.0
Middle Income Countries						
Brazil	10.9	11.9	12.6	12.6	11.2	13.0
Mexico	29.9	31.2	32.6	31.7	32.3	35.3
South Korea	49.4	55.7	56.3	53.9	50.3	45.9

Table 1.2 The Extent of Globalization (exports/GDP) (Source: http://databank.worldbank.org/data/)

Country	2010	2011	2012	2013	2014	2015
Lower Income Countries						
Chad	36.8	38.9	36.9	32.2	34.2	29.8
China	29.4	28.5	27.3	26.4	23.9	22.4
India	22.0	23.9	24.0	24.8	22.9	-
Nigeria	25.3	31.3	31.4	18.0	18.4	-

Table 1.2 The Extent of Globalization (exports/GDP) (Source: http://databank.worldbank.org/data/)

In recent decades, the export/GDP ratio has generally risen, both worldwide and for the U.S. economy. Interestingly, the share of U.S. exports in proportion to the U.S. economy is well below the global average, in part because large economies like the United States can contain more of the division of labor inside their national borders. However, smaller economies like Belgium, Korea, and Canada need to trade across their borders with other countries to take full advantage of division of labor, specialization, and economies of scale. In this sense, the enormous U.S. economy is less affected by globalization than most other countries.

Table 1.2 indicates that many medium and low income countries around the world, like Mexico and China, have also experienced a surge of globalization in recent decades. If an astronaut in orbit could put on special glasses that make all economic transactions visible as brightly colored lines and look down at Earth, the astronaut would see the planet covered with connections.

Despite the rise in globalization over the last few decades, in recent years we've seen significant pushback against globalization from people across the world concerned about loss of jobs, loss of political sovereignty, and increased economic inequality. Prominent examples of this pushback include the 2016 vote in Great Britain to exit the European Union (i.e. Brexit), and the election of Donald J. Trump for President of the United States.

Hopefully, you now have an idea about economics. Before you move to any other chapter of study, be sure to read the very important appendix to this chapter called **The Use of Mathematics in Principles of Economics**. It is essential that you learn more about how to read and use models in economics.

Bring it Home

Decisions ... Decisions in the Social Media Age

The world we live in today provides nearly instant access to a wealth of information. Consider that as recently as the late 1970s, the *Farmer's Almanac*, along with the Weather Bureau of the U.S. Department of Agriculture, were the primary sources American farmers used to determine when to plant and harvest their crops. Today, farmers are more likely to access, online, weather forecasts from the National Oceanic and Atmospheric Administration or watch the Weather Channel. After all, knowing the upcoming forecast could drive when to harvest crops. Consequently, knowing the upcoming weather could change the amount of crop harvested.

Some relatively new information forums, such as Facebook, are rapidly changing how information is distributed; hence, influencing decision making. In 2014, the Pew Research Center reported that 71% of online adults use Facebook. This social media forum posts topics ranging from the National Basketball Association, to celebrity singers and performers, to farmers.

Information helps us make decisions as simple as what to wear today to how many reporters the media should send to cover a crash. Each of these decisions is an economic decision. After all, resources are scarce. If the media send ten reporters to cover an accident, they are not available to cover other stories or complete other tasks. Information provides the necessary knowledge to make the best possible decisions on how to utilize scarce resources. Welcome to the world of economics!

KEY TERMS

circular flow diagram a diagram that views the economy as consisting of households and firms interacting in a goods and services market and a labor market

command economy an economy where economic decisions are passed down from government authority and where the government owns the resources

division of labor the way in which different workers divide required tasks to produce a good or service

economics the study of how humans make choices under conditions of scarcity

economies of scale when the average cost of producing each individual unit declines as total output increases

exports products (goods and services) made domestically and sold abroad

fiscal policy economic policies that involve government spending and taxes

globalization the trend in which buying and selling in markets have increasingly crossed national borders

goods and services market a market in which firms are sellers of what they produce and households are buyers

gross domestic product (GDP) measure of the size of total production in an economy

imports products (goods and services) made abroad and then sold domestically

labor market the market in which households sell their labor as workers to business firms or other employers

macroeconomics the branch of economics that focuses on broad issues such as growth, unemployment, inflation, and trade balance

market interaction between potential buyers and sellers; a combination of demand and supply

market economy an economy where economic decisions are decentralized, private individuals own resources, and businesses supply goods and services based on demand

microeconomics the branch of economics that focuses on actions of particular agents within the economy, like households, workers, and business firms

model see theory

monetary policy policy that involves altering the level of interest rates, the availability of credit in the economy, and the extent of borrowing

private enterprise system where private individuals or groups of private individuals own and operate the means of production (resources and businesses)

scarcity when human wants for goods and services exceed the available supply

specialization when workers or firms focus on particular tasks for which they are well-suited within the overall production process

theory a representation of an object or situation that is simplified while including enough of the key features to help us understand the object or situation

traditional economy typically an agricultural economy where things are done the same as they have always been done

underground economy a market where the buyers and sellers make transactions in violation of one or more

government regulations

KEY CONCEPTS AND SUMMARY

1.1 What Is Economics, and Why Is It Important?

Economics seeks to solve the problem of scarcity, which is when human wants for goods and services exceed the available supply. A modern economy displays a division of labor, in which people earn income by specializing in what they produce and then use that income to purchase the products they need or want. The division of labor allows individuals and firms to specialize and to produce more for several reasons: a) It allows the agents to focus on areas of advantage due to natural factors and skill levels; b) It encourages the agents to learn and invent; c) It allows agents to take advantage of economies of scale. Division and specialization of labor only work when individuals can purchase what they do not produce in markets. Learning about economics helps you understand the major problems facing the world today, prepares you to be a good citizen, and helps you become a well-rounded thinker.

1.2 Microeconomics and Macroeconomics

Microeconomics and macroeconomics are two different perspectives on the economy. The microeconomic perspective focuses on parts of the economy: individuals, firms, and industries. The macroeconomic perspective looks at the economy as a whole, focusing on goals like growth in the standard of living, unemployment, and inflation. Macroeconomics has two types of policies for pursuing these goals: monetary policy and fiscal policy.

1.3 How Economists Use Theories and Models to Understand Economic Issues

Economists analyze problems differently than do other disciplinary experts. The main tools economists use are economic theories or models. A theory is not an illustration of the answer to a problem. Rather, a theory is a tool for determining the answer.

1.4 How To Organize Economies: An Overview of Economic Systems

We can organize societies as traditional, command, or market-oriented economies. Most societies are a mix. The last few decades have seen globalization evolve as a result of growth in commercial and financial networks that cross national borders, making businesses and workers from different economies increasingly interdependent.

SELF-CHECK QUESTIONS

1. What is scarcity? Can you think of two causes of scarcity?

2. Residents of the town of Smithfield like to consume hams, but each ham requires 10 people to produce it and takes a month. If the town has a total of 100 people, what is the maximum amount of ham the residents can consume in a month?

3. A consultant works for $200 per hour. She likes to eat vegetables, but is not very good at growing them. Why does it make more economic sense for her to spend her time at the consulting job and shop for her vegetables?

4. A computer systems engineer could paint his house, but it makes more sense for him to hire a painter to do it. Explain why.

5. What would be another example of a “system” in the real world that could serve as a metaphor for micro and macroeconomics?

6. Suppose we extend the circular flow model to add imports and exports. Copy the circular flow diagram onto a sheet of paper and then add a foreign country as a third agent. Draw a rough sketch of the flows of imports, exports, and the payments for each on your diagram.

7. What is an example of a problem in the world today, not mentioned in the chapter, that has an economic dimension?

8. The chapter defines *private enterprise* as a characteristic of market-oriented economies. What would *public enterprise* be? *Hint*: It is a characteristic of command economies.

9. Why might Belgium, France, Italy, and Sweden have a higher export to GDP ratio than the United States?

REVIEW QUESTIONS

10. Give the three reasons that explain why the division of labor increases an economy's level of production.

11. What are three reasons to study economics?

12. What is the difference between microeconomics and macroeconomics?

13. What are examples of individual economic agents?

14. What are the three main goals of macroeconomics?

15. How did John Maynard Keynes define economics?

16. Are households primarily buyers or sellers in the goods and services market? In the labor market?

17. Are firms primarily buyers or sellers in the goods and services market? In the labor market?

18. What are the three ways that societies can organize themselves economically?

19. What is globalization? How do you think it might have affected the economy over the past decade?

CRITICAL THINKING QUESTIONS

20. Suppose you have a team of two workers: one is a baker and one is a chef. Explain why the kitchen can produce more meals in a given period of time if each worker specializes in what they do best than if each worker tries to do everything from appetizer to dessert.

21. Why would division of labor without trade not work?

22. Can you think of any examples of *free* goods, that is, goods or services that are not scarce?

23. A balanced federal budget and a balance of trade are secondary goals of macroeconomics, while growth in the standard of living (for example) is a primary goal. Why do you think that is so?

24. Macroeconomics is an aggregate of what happens at the microeconomic level. Would it be possible for what happens at the macro level to differ from how economic agents would react to some stimulus at the micro level? *Hint:* Think about the behavior of crowds.

25. Why is it unfair or meaningless to criticize a theory as "unrealistic?"

26. Suppose, as an economist, you are asked to analyze an issue unlike anything you have ever done before. Also, suppose you do not have a specific model for analyzing that issue. What should you do? *Hint:* What would a carpenter do in a similar situation?

27. Why do you think that most modern countries' economies are a mix of command and market types?

28. Can you think of ways that globalization has helped you economically? Can you think of ways that it has not?

2 | Choice in a World of Scarcity

Figure 2.1 Choices and Tradeoffs In general, the higher the degree, the higher the salary, so why aren't more people pursuing higher degrees? The short answer: choices and tradeoffs. (Credit: modification of work by "Jim, the Photographer"/Flickr Creative Commons)

Bring it Home

Choices ... To What Degree?

In 2015, the median income for workers who hold master's degrees varies from males to females. The average of the two is $2,951 weekly. Multiply this average by 52 weeks, and you get an average salary of $153,452. Compare that to the median weekly earnings for a full-time worker over 25 with no higher than a bachelor's degree: $1,224 weekly and $63,648 a year. What about those with no higher than a high school diploma in 2015? They earn just $664 weekly and $34,528 over 12 months. In other words, says the Bureau of Labor Statistics (BLS), earning a bachelor's degree boosted salaries 54% over what you would have earned if you had stopped your education after high school. A master's degree yields a salary almost double that of a high school diploma.

Given these statistics, we might expect many people to choose to go to college and at least earn a bachelor's degree. Assuming that people want to improve their material well-being, it seems like they would make those choices that provide them with the greatest opportunity to consume goods and services. As it turns out, the analysis is not nearly as simple as this. In fact, in 2014, the BLS reported that while almost 88% of the population in the United States had a high school diploma, only 33.6% of 25–65 year olds had bachelor's degrees, and only 7.4% of 25–65 year olds in 2014 had earned a master's.

This brings us to the subject of this chapter: why people make the choices they make and how economists explain those choices.

Introduction to Choice in a World of Scarcity

In this chapter, you will learn about:

- How Individuals Make Choices Based on Their Budget Constraint
- The Production Possibilities Frontier and Social Choices
- Confronting Objections to the Economic Approach

You will learn quickly when you examine the relationship between economics and scarcity that choices involve tradeoffs. Every choice has a cost.

In 1968, the Rolling Stones recorded "You Can't Always Get What You Want." Economists chuckled, because they had been singing a similar tune for decades. English economist Lionel Robbins (1898–1984), in his *Essay on the Nature and Significance of Economic Science* in 1932, described not always getting what you want in this way:

> The time at our disposal is limited. There are only twenty-four hours in the day. We have to choose between the different uses to which they may be put. ... Everywhere we turn, if we choose one thing we must relinquish others which, in different circumstances, we would wish not to have relinquished. Scarcity of means to satisfy given ends is an almost ubiquitous condition of human nature.

Because people live in a world of scarcity, they cannot have all the time, money, possessions, and experiences they wish. Neither can society.

This chapter will continue our discussion of scarcity and the economic way of thinking by first introducing three critical concepts: opportunity cost, marginal decision making, and diminishing returns. Later, it will consider whether the economic way of thinking accurately describes either how we *make* choices and how we *should* make them.

2.1 | How Individuals Make Choices Based on Their Budget Constraint

By the end of this section, you will be able to:

- Calculate and graph budget constraints
- Explain opportunity sets and opportunity costs
- Evaluate the law of diminishing marginal utility
- Explain how marginal analysis and utility influence choices

Consider the typical consumer's budget problem. Consumers have a limited amount of income to spend on the things they need and want. Suppose Alphonso has $10 in spending money each week that he can allocate between bus tickets for getting to work and the burgers that he eats for lunch. Burgers cost $2 each, and bus tickets are 50 cents each. We can see Alphonso's budget problem in **Figure 2.2**.

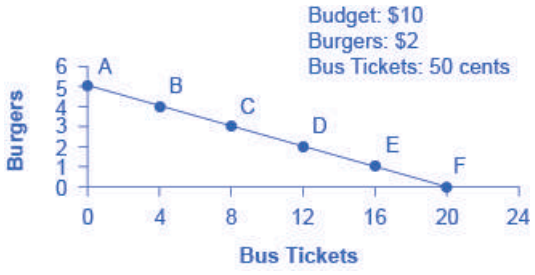


Figure 2.2 The Budget Constraint: Alphonso's Consumption Choice Opportunity Frontier Each point on the budget constraint represents a combination of burgers and bus tickets whose total cost adds up to Alphonso's budget of $10. The relative price of burgers and bus tickets determines the slope of the budget constraint. All along the budget set, giving up one burger means gaining four bus tickets.

The vertical axis in the figure shows burger purchases and the horizontal axis shows bus ticket purchases. If Alphonso spends all his money on burgers, he can afford five per week. ($10 per week/$2 per burger = 5 burgers per week.) However, if he does this, he will not be able to afford any bus tickets. Point A in the figure shows the choice (zero bus tickets and five burgers). Alternatively, if Alphonso spends all his money on bus tickets, he can afford 20 per week. ($10 per week/$0.50 per bus ticket = 20 bus tickets per week.) Then, however, he will not be able to afford any burgers. Point F shows this alternative choice (20 bus tickets and zero burgers).

If we connect all the points between A and F, we get Alphonso's **budget constraint**. This indicates all the combination of burgers and bus tickets Alphonso can afford, given the price of the two goods and his budget amount.

If Alphonso is like most people, he will choose some combination that includes both bus tickets and burgers. That is, he will choose some combination on the budget constraint that is between points A and F. Every point on (or inside) the constraint shows a combination of burgers and bus tickets that Alphonso can afford. Any point outside the constraint is not affordable, because it would cost more money than Alphonso has in his budget.

The budget constraint clearly shows the tradeoff Alphonso faces in choosing between burgers and bus tickets. Suppose he is currently at point D, where he can afford 12 bus tickets and two burgers. What would it cost Alphonso for one more burger? It would be natural to answer $2, but that's not the way economists think. Instead they ask, how many bus tickets would Alphonso have to give up to get one more burger, while staying within his budget? Since bus tickets cost 50 cents, Alphonso would have to give up four to afford one more burger. That is the true cost to Alphonso.

The Concept of Opportunity Cost

Economists use the term **opportunity cost** to indicate what one must give up to obtain what he or she desires. The idea behind opportunity cost is that the cost of one item is the lost opportunity to do or consume something else. In short, opportunity cost is the value of the next best alternative. For Alphonso, the opportunity cost of a burger is the four bus tickets he would have to give up. He would decide whether or not to choose the burger depending on whether the value of the burger exceeds the value of the forgone alternative—in this case, bus tickets. Since people must choose, they inevitably face tradeoffs in which they have to give up things they desire to obtain other things they desire more.

Link It Up

View this website (http://openstaxcollege.org/l/linestanding) for an example of opportunity cost—paying someone else to wait in line for you.

A fundamental principle of economics is that every choice has an opportunity cost. If you sleep through your economics class, the opportunity cost is the learning you miss from not attending class. If you spend your income on video games, you cannot spend it on movies. If you choose to marry one person, you give up the opportunity to marry anyone else. In short, opportunity cost is all around us and part of human existence.

The following Work It Out feature shows a step-by-step analysis of a budget constraint calculation. Read through it to understand another important concept—slope—that we further explain in the appendix **The Use of Mathematics in Principles of Economics**.

Work It Out

Understanding Budget Constraints

Budget constraints are easy to understand if you apply a little math. The appendix **The Use of Mathematics in Principles of Economics** explains all the math you are likely to need in this book. Therefore, if math is not your strength, you might want to take a look at the appendix.

Step 1: The equation for any budget constraint is:

$$\text{Budget} = P_1 \times Q_1 + P_2 \times Q_2$$

where P and Q are the price and quantity of items purchased (which we assume here to be two items) and Budget is the amount of income one has to spend.

Step 2. Apply the budget constraint equation to the scenario. In Alphonso's case, this works out to be:

$$\begin{aligned} \text{Budget} &= P_1 \times Q_1 + P_2 \times Q_2 \\ \$10 \text{ budget} &= \$2 \text{ per burger} \times \text{quantity of burgers} + \$0.50 \text{ per bus ticket} \times \text{quantity of bus tickets} \\ \$10 &= \$2 \times Q_{\text{burgers}} + \$0.50 \times Q_{\text{bus tickets}} \end{aligned}$$

Step 3. Using a little algebra, we can turn this into the familiar equation of a line:

$$y = b + mx$$

For Alphonso, this is:

$$\$10 = \$2 \times Q_{\text{burgers}} + \$0.50 \times Q_{\text{bus tickets}}$$

Step 4. Simplify the equation. Begin by multiplying both sides of the equation by 2:

$$\begin{aligned} 2 \times 10 &= 2 \times 2 \times Q_{\text{burgers}} + 2 \times 0.5 \times Q_{\text{bus tickets}} \\ 20 &= 4 \times Q_{\text{burgers}} + 1 \times Q_{\text{bus tickets}} \end{aligned}$$

Step 5. Subtract one bus ticket from both sides:

$$20 - Q_{\text{bus tickets}} = 4 \times Q_{\text{burgers}}$$

Divide each side by 4 to yield the answer:

$$\begin{aligned} 5 - 0.25 \times Q_{\text{bus tickets}} &= Q_{\text{burgers}} \\ \text{or} \\ Q_{\text{burgers}} &= 5 - 0.25 \times Q_{\text{bus tickets}} \end{aligned}$$

Step 6. Notice that this equation fits the budget constraint in **Figure 2.2**. The vertical intercept is 5 and the slope is –0.25, just as the equation says. If you plug 20 bus tickets into the equation, you get 0 burgers. If you plug other numbers of bus tickets into the equation, you get the results (see **Table 2.1**), which are the points on Alphonso's budget constraint.

Point	Quantity of Burgers (at $2)	Quantity of Bus Tickets (at 50 cents)
A	5	0
B	4	4
C	3	8
D	2	12

Table 2.1

Point	Quantity of Burgers (at $2)	Quantity of Bus Tickets (at 50 cents)
E	1	16
F	0	20

Table 2.1

Step 7. Notice that the slope of a budget constraint always shows the opportunity cost of the good which is on the horizontal axis. For Alphonso, the slope is −0.25, indicating that for every bus ticket he buys, he must give up 1/4 burger. To phrase it differently, for every four tickets he buys, Alphonso must give up 1 burger.

There are two important observations here. First, the algebraic sign of the slope is negative, which means that the only way to get more of one good is to give up some of the other. Second, we define the slope as the price of bus tickets (whatever is on the horizontal axis in the graph) divided by the price of burgers (whatever is on the vertical axis), in this case $0.50/$2 = 0.25. If you want to determine the opportunity cost quickly, just divide the two prices.

Identifying Opportunity Cost

In many cases, it is reasonable to refer to the opportunity cost as the price. If your cousin buys a new bicycle for $300, then $300 measures the amount of "other consumption" that he has forsaken. For practical purposes, there may be no special need to identify the specific alternative product or products that he could have bought with that $300, but sometimes the price as measured in dollars may not accurately capture the true opportunity cost. This problem can loom especially large when costs of time are involved.

For example, consider a boss who decides that all employees will attend a two-day retreat to "build team spirit." The out-of-pocket monetary cost of the event may involve hiring an outside consulting firm to run the retreat, as well as room and board for all participants. However, an opportunity cost exists as well: during the two days of the retreat, none of the employees are doing any other work.

Attending college is another case where the opportunity cost exceeds the monetary cost. The out-of-pocket costs of attending college include tuition, books, room and board, and other expenses. However, in addition, during the hours that you are attending class and studying, it is impossible to work at a paying job. Thus, college imposes both an out-of-pocket cost and an opportunity cost of lost earnings.

Clear It Up

What is the opportunity cost associated with increased airport security measures?

After the terrorist plane hijackings on September 11, 2001, many steps were proposed to improve air travel safety. For example, the federal government could provide armed "sky marshals" who would travel inconspicuously with the rest of the passengers. The cost of having a sky marshal on every flight would be roughly $3 billion per year. Retrofitting all U.S. planes with reinforced cockpit doors to make it harder for terrorists to take over the plane would have a price tag of $450 million. Buying more sophisticated security equipment for airports, like three-dimensional baggage scanners and cameras linked to face recognition software, could cost another $2 billion.

However, the single biggest cost of greater airline security does not involve spending money. It is the opportunity cost of additional waiting time at the airport. According to the United States Department of Transportation (DOT), there were 895.5 million systemwide (domestic and international) scheduled service passengers in 2015. Since the 9/11 hijackings, security screening has become more intensive, and consequently, the procedure takes longer than in the past. Say that, on average, each air passenger spends

an extra 30 minutes in the airport per trip. Economists commonly place a value on time to convert an opportunity cost in time into a monetary figure. Because many air travelers are relatively high-paid business people, conservative estimates set the average price of time for air travelers at $20 per hour. By these back-of-the-envelope calculations, the opportunity cost of delays in airports could be as much as 800 million × 0.5 hours × $20/hour, or $8 billion per year. Clearly, the opportunity costs of waiting time can be just as important as costs that involve direct spending.

In some cases, realizing the opportunity cost can alter behavior. Imagine, for example, that you spend $8 on lunch every day at work. You may know perfectly well that bringing a lunch from home would cost only $3 a day, so the opportunity cost of buying lunch at the restaurant is $5 each day (that is, the $8 buying lunch costs minus the $3 your lunch from home would cost). Five dollars each day does not seem to be that much. However, if you project what that adds up to in a year—250 days a year × $5 per day equals $1,250, the cost, perhaps, of a decent vacation. If you describe the opportunity cost as "a nice vacation" instead of "$5 a day," you might make different choices.

Marginal Decision-Making and Diminishing Marginal Utility

The budget constraint framework helps to emphasize that most choices in the real world are not about getting all of one thing or all of another; that is, they are not about choosing either the point at one end of the budget constraint or else the point all the way at the other end. Instead, most choices involve **marginal analysis**, which means examining the benefits and costs of choosing a little more or a little less of a good. People naturally compare costs and benefits, but often we look at total costs and total benefits, when the optimal choice necessitates comparing how costs and benefits change from one option to another. You might think of marginal analysis as "change analysis." Marginal analysis is used throughout economics.

We now turn to the notion of **utility**. People desire goods and services for the satisfaction or utility those goods and services provide. Utility, as we will see in the chapter on **Consumer Choices**, is subjective but that does not make it less real. Economists typically assume that the more of some good one consumes (for example, slices of pizza), the more utility one obtains. At the same time, the utility a person receives from consuming the first unit of a good is typically more than the utility received from consuming the fifth or the tenth unit of that same good. When Alphonso chooses between burgers and bus tickets, for example, the first few bus rides that he chooses might provide him with a great deal of utility—perhaps they help him get to a job interview or a doctor's appointment. However, later bus rides might provide much less utility—they may only serve to kill time on a rainy day. Similarly, the first burger that Alphonso chooses to buy may be on a day when he missed breakfast and is ravenously hungry. However, if Alphonso has a burger every single day, the last few burgers may taste pretty boring. The general pattern that consumption of the first few units of any good tends to bring a higher level of utility to a person than consumption of later units is a common pattern. Economists refer to this pattern as the **law of diminishing marginal utility**, which means that as a person receives more of a good, the additional (or marginal) utility from each additional unit of the good declines. In other words, the first slice of pizza brings more satisfaction than the sixth.

The law of diminishing marginal utility explains why people and societies rarely make all-or-nothing choices. You would not say, "My favorite food is ice cream, so I will eat nothing but ice cream from now on." Instead, even if you get a very high level of utility from your favorite food, if you ate it exclusively, the additional or marginal utility from those last few servings would not be very high. Similarly, most workers do not say: "I enjoy leisure, so I'll never work." Instead, workers recognize that even though some leisure is very nice, a combination of all leisure and no income is not so attractive. The budget constraint framework suggests that when people make choices in a world of scarcity, they will use marginal analysis and think about whether they would prefer a little more or a little less.

A rational consumer would only purchase additional units of some product as long as the marginal utility exceeds the opportunity cost. Suppose Alphonso moves down his budget constraint from Point A to Point B to Point C and further. As he consumes more bus tickets, the marginal utility of bus tickets will diminish, while the opportunity cost, that is, the marginal utility of foregone burgers, will increase. Eventually, the opportunity cost will exceed the marginal utility of an additional bus ticket. If Alphonso is rational, he won't purchase more bus tickets once the marginal utility just equals the opportunity cost. While we can't (yet) say exactly how many bus tickets Alphonso will buy, that number is unlikely to be the most he can afford, 20.

Sunk Costs

In the budget constraint framework, all decisions involve what will happen next: that is, what quantities of goods will

you consume, how many hours will you work, or how much will you save. These decisions do not look back to past choices. Thus, the budget constraint framework assumes that **sunk costs**, which are costs that were incurred in the past and cannot be recovered, should not affect the current decision.

Consider the case of Selena, who pays $8 to see a movie, but after watching the film for 30 minutes, she knows that it is truly terrible. Should she stay and watch the rest of the movie because she paid for the ticket, or should she leave? The money she spent is a sunk cost, and unless the theater manager is sympathetic, Selena will not get a refund. However, staying in the movie still means paying an opportunity cost in time. Her choice is whether to spend the next 90 minutes suffering through a cinematic disaster or to do something—anything—else. The lesson of sunk costs is to forget about the money and time that is irretrievably gone and instead to focus on the marginal costs and benefits of current and future options.

For people and firms alike, dealing with sunk costs can be frustrating. It often means admitting an earlier error in judgment. Many firms, for example, find it hard to give up on a new product that is doing poorly because they spent so much money in creating and launching the product. However, the lesson of sunk costs is to ignore them and make decisions based on what will happen in the future.

From a Model with Two Goods to One of Many Goods

The budget constraint diagram containing just two goods, like most models used in this book, is not realistic. After all, in a modern economy people choose from thousands of goods. However, thinking about a model with many goods is a straightforward extension of what we discussed here. Instead of drawing just one budget constraint, showing the tradeoff between two goods, you can draw multiple budget constraints, showing the possible tradeoffs between many different pairs of goods. In more advanced classes in economics, you would use mathematical equations that include many possible goods and services that can be purchased, together with their quantities and prices, and show how the total spending on all goods and services is limited to the overall budget available. The graph with two goods that we presented here clearly illustrates that every choice has an opportunity cost, which is the point that does carry over to the real world.

2.2 | The Production Possibilities Frontier and Social Choices

By the end of this section, you will be able to:

- Interpret production possibilities frontier graphs
- Contrast a budget constraint and a production possibilities frontier
- Explain the relationship between a production possibilities frontier and the law of diminishing returns
- Contrast productive efficiency and allocative efficiency
- Define comparative advantage

Just as individuals cannot have everything they want and must instead make choices, society as a whole cannot have everything it might want, either. This section of the chapter will explain the constraints society faces, using a model called the **production possibilities frontier (PPF)**. There are more similarities than differences between individual choice and social choice. As you read this section, focus on the similarities.

Because society has limited resources (e.g., labor, land, capital, raw materials) at any point in time, there is a limit to the quantities of goods and services it can produce. Suppose a society desires two products, healthcare and education. The production possibilities frontier in **Figure 2.3** illustrates this situation.

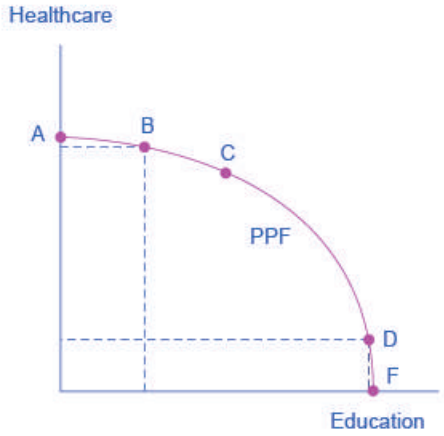


Figure 2.3 A Healthcare vs. Education Production Possibilities Frontier This production possibilities frontier shows a tradeoff between devoting social resources to healthcare and devoting them to education. At A all resources go to healthcare and at B, most go to healthcare. At D most resources go to education, and at F, all go to education.

Figure 2.3 shows healthcare on the vertical axis and education on the horizontal axis. If the society were to allocate all of its resources to healthcare, it could produce at point A. However, it would not have any resources to produce education. If it were to allocate all of its resources to education, it could produce at point F. Alternatively, the society could choose to produce any combination of healthcare and education on the production possibilities frontier. In effect, the production possibilities frontier plays the same role for society as the budget constraint plays for Alphonso. Society can choose any combination of the two goods on or inside the PPF. However, it does not have enough resources to produce outside the PPF.

Most importantly, the production possibilities frontier clearly shows the tradeoff between healthcare and education. Suppose society has chosen to operate at point B, and it is considering producing more education. Because the PPF is downward sloping from left to right, the only way society can obtain more education is by giving up some healthcare. That is the tradeoff society faces. Suppose it considers moving from point B to point C. What would the opportunity cost be for the additional education? The opportunity cost would be the healthcare society has to forgo. Just as with Alphonso's budget constraint, the **slope** of the production possibilities frontier shows the opportunity cost. By now you might be saying, "Hey, this PPF is sounding like the budget constraint." If so, read the following Clear It Up feature.

Clear It Up

What's the difference between a budget constraint and a PPF?

There are two major differences between a budget constraint and a production possibilities frontier. The first is the fact that the budget constraint is a straight line. This is because its slope is given by the relative prices of the two goods, which from the point of view of an individual consumer, are fixed, so the slope doesn't change. In contrast, the PPF has a curved shape because of the law of the diminishing returns. Thus, the slope is different at various points on the PPF. The second major difference is the absence of specific numbers on the axes of the PPF. There are no specific numbers because we do not know the exact amount of resources this imaginary economy has, nor do we know how many resources it takes to produce healthcare and how many resources it takes to produce education. If this were a real world example, that data would be available.

Whether or not we have specific numbers, conceptually we can measure the opportunity cost of additional education as society moves from point B to point C on the PPF. We measure the additional education by the horizontal distance between B and C. The foregone healthcare is given by the vertical distance between B and C. The slope of the PPF between B and C is (approximately) the vertical distance (the "rise") over the horizontal distance (the "run"). This is the opportunity cost of the additional education.

The Shape of the PPF and the Law of Diminishing Returns

The budget constraints that we presented earlier in this chapter, showing individual choices about what quantities of goods to consume, were all straight lines. The reason for these straight lines was that the relative prices of the two goods in the **consumption budget constraint** determined the slope of the budget constraint. However, we drew the production possibilities frontier for healthcare and education as a curved line. Why does the PPF have a different shape?

To understand why the PPF is curved, start by considering point A at the top left-hand side of the PPF. At point A, all available resources are devoted to healthcare and none are left for education. This situation would be extreme and even ridiculous. For example, children are seeing a doctor every day, whether they are sick or not, but not attending school. People are having cosmetic surgery on every part of their bodies, but no high school or college education exists. Now imagine that some of these resources are diverted from healthcare to education, so that the economy is at point B instead of point A. Diverting some resources away from A to B causes relatively little reduction in health because the last few marginal dollars going into healthcare services are not producing much additional gain in health. However, putting those marginal dollars into education, which is completely without resources at point A, can produce relatively large gains. For this reason, the shape of the PPF from A to B is relatively flat, representing a relatively small drop-off in health and a relatively large gain in education.

Now consider the other end, at the lower right, of the production possibilities frontier. Imagine that society starts at choice D, which is devoting nearly all resources to education and very few to healthcare, and moves to point F, which is devoting *all* spending to education and none to healthcare. For the sake of concreteness, you can imagine that in the movement from D to F, the last few doctors must become high school science teachers, the last few nurses must become school librarians rather than dispensers of vaccinations, and the last few emergency rooms are turned into kindergartens. The gains to education from adding these last few resources to education are very small. However, the opportunity cost lost to health will be fairly large, and thus the slope of the PPF between D and F is steep, showing a large drop in health for only a small gain in education.

The lesson is not that society is likely to make an extreme choice like devoting no resources to education at point A or no resources to health at point F. Instead, the lesson is that the gains from committing additional marginal resources to education depend on how much is already being spent. If on the one hand, very few resources are currently committed to education, then an increase in resources used can bring relatively large gains. On the other hand, if a large number of resources are already committed to education, then committing additional resources will bring relatively smaller gains.

This pattern is common enough that economists have given it a name: the **law of diminishing returns**, which holds that as additional increments of resources are added to a certain purpose, the marginal benefit from those additional increments will decline. (The law of diminishing marginal utility that we introduced in the last section is a more specific case of the law of diminishing returns.) When government spends a certain amount more on reducing crime, for example, the original gains in reducing crime could be relatively large. However, additional increases typically cause relatively smaller reductions in crime, and paying for enough police and security to reduce crime to nothing at all would be tremendously expensive.

The curvature of the production possibilities frontier shows that as we add more resources to education, moving from left to right along the horizontal axis, the original gains are fairly large, but gradually diminish. Thus, the slope of the PPF is relatively flat. By contrast, as we add more resources to healthcare, moving from bottom to top on the vertical axis, the original gains are fairly large, but again gradually diminish. Thus, the slope of the PPF is relatively steep. In this way, the law of diminishing returns produces the outward-bending shape of the production possibilities frontier.

Productive Efficiency and Allocative Efficiency

The study of economics does not presume to tell a society what choice it should make along its production possibilities frontier. In a market-oriented economy with a democratic government, the choice will involve a mixture of decisions by individuals, firms, and government. However, economics can point out that some choices are unambiguously better than others. This observation is based on the concept of efficiency. In everyday usage, efficiency refers to lack of waste. An inefficient machine operates at high cost, while an efficient machine operates at lower cost, because it is not wasting energy or materials. An inefficient organization operates with long delays and high costs, while an efficient organization meets schedules, is focused, and performs within budget.

The production possibilities frontier can illustrate two kinds of efficiency: productive efficiency and allocative efficiency. **Figure 2.4** illustrates these ideas using a production possibilities frontier between healthcare and

education.

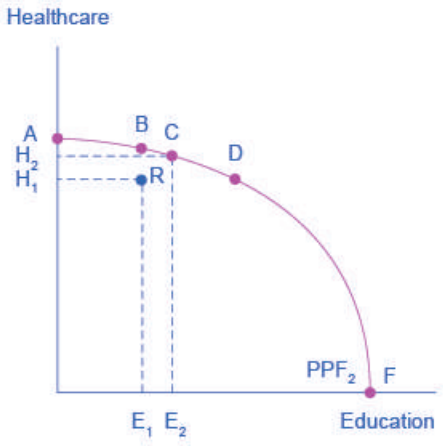


Figure 2.4 Productive and Allocative Efficiency Productive efficiency means it is impossible to produce more of one good without decreasing the quantity that is produced of another good. Thus, all choices along a given PPF like B, C, and D display productive efficiency, but R does not. Allocative efficiency means that the particular mix of goods being produced—that is, the specific choice along the production possibilities frontier—represents the allocation that society most desires.

Productive efficiency means that, given the available inputs and technology, it is impossible to produce more of one good without decreasing the quantity that is produced of another good. All choices on the PPF in **Figure 2.4**, including A, B, C, D, and F, display productive efficiency. As a firm moves from any one of these choices to any other, either healthcare increases and education decreases or vice versa. However, any choice inside the production possibilities frontier is productively inefficient and wasteful because it is possible to produce more of one good, the other good, or some combination of both goods.

For example, point R is productively inefficient because it is possible at choice C to have more of both goods: education on the horizontal axis is higher at point C than point R (E_2 is greater than E_1), and healthcare on the vertical axis is also higher at point C than point R (H_2 is great than H_1).

We can show the particular mix of goods and services produced—that is, the specific combination of selected healthcare and education along the production possibilities frontier—as a ray (line) from the origin to a specific point on the PPF. Output mixes that had more healthcare (and less education) would have a steeper ray, while those with more education (and less healthcare) would have a flatter ray.

Allocative efficiency means that the particular combination of goods and services on the production possibility curve that a society produces represents the combination that society most desires. How to determine what a society desires can be a controversial question, and is usually a discussion in political science, sociology, and philosophy classes as well as in economics. At its most basic, allocative efficiency means producers supply the quantity of each product that consumers demand. Only one of the productively efficient choices will be the allocatively efficient choice for society as a whole.

Why Society Must Choose

In **Welcome to Economics!** we learned that every society faces the problem of scarcity, where limited resources conflict with unlimited needs and wants. The production possibilities curve illustrates the choices involved in this dilemma.

Every economy faces two situations in which it may be able to expand consumption of all goods. In the first case, a society may discover that it has been using its resources inefficiently, in which case by improving efficiency and producing on the production possibilities frontier, it can have more of all goods (or at least more of some and less of none). In the second case, as resources grow over a period of years (e.g., more labor and more capital), the economy grows. As it does, the production possibilities frontier for a society will tend to shift outward and society will be able to afford more of all goods.

However, improvements in productive efficiency take time to discover and implement, and economic growth happens only gradually. Thus, a society must choose between tradeoffs in the present. For government, this process often involves trying to identify where additional spending could do the most good and where reductions in spending would do the least harm. At the individual and firm level, the market economy coordinates a process in which firms seek to produce goods and services in the quantity, quality, and price that people want. However, for both the government and the market economy in the short term, increases in production of one good typically mean offsetting decreases somewhere else in the economy.

The PPF and Comparative Advantage

While every society must choose how much of each good or service it should produce, it does not need to produce every single good it consumes. Often how much of a good a country decides to produce depends on how expensive it is to produce it versus buying it from a different country. As we saw earlier, the curvature of a country's PPF gives us information about the tradeoff between devoting resources to producing one good versus another. In particular, its slope gives the opportunity cost of producing one more unit of the good in the x-axis in terms of the other good (in the y-axis). Countries tend to have different opportunity costs of producing a specific good, either because of different climates, geography, technology, or skills.

Suppose two countries, the US and Brazil, need to decide how much they will produce of two crops: sugar cane and wheat. Due to its climatic conditions, Brazil can produce quite a bit of sugar cane per acre but not much wheat. Conversely, the U.S. can produce large amounts of wheat per acre, but not much sugar cane. Clearly, Brazil has a lower opportunity cost of producing sugar cane (in terms of wheat) than the U.S. The reverse is also true: the U.S. has a lower opportunity cost of producing wheat than Brazil. We illustrate this by the PPFs of the two countries in **Figure 2.5**.

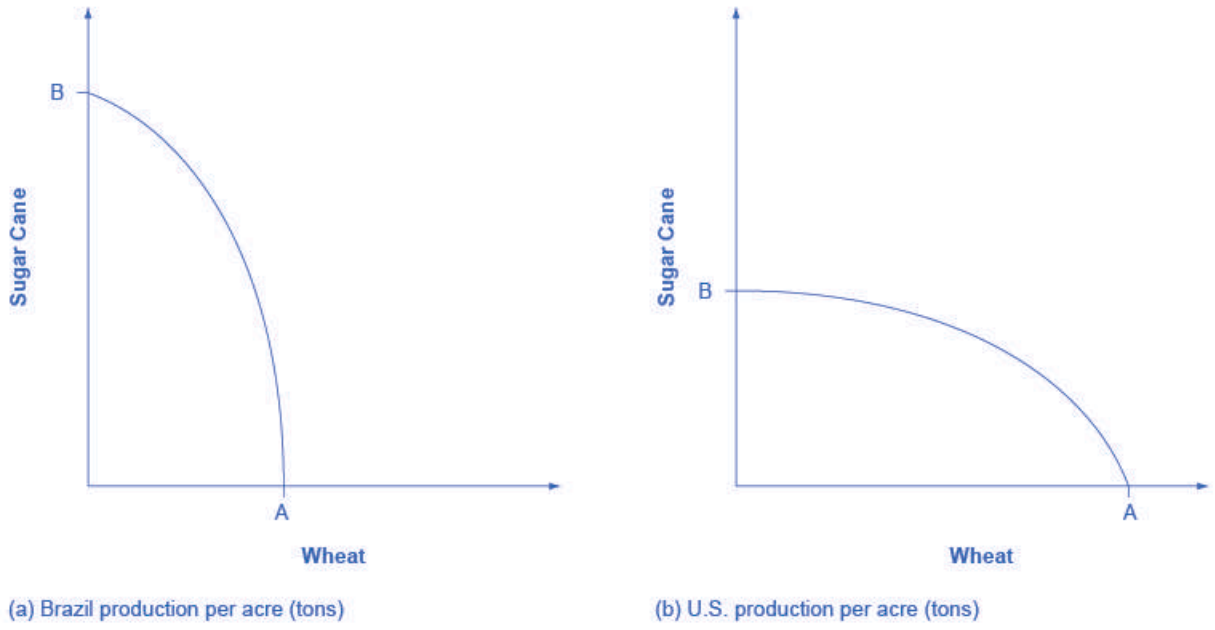


Figure 2.5 Production Possibility Frontier for the U.S. and Brazil The U.S. PPF is flatter than the Brazil PPF implying that the opportunity cost of wheat in terms of sugar cane is lower in the U.S. than in Brazil. Conversely, the opportunity cost of sugar cane is lower in Brazil. The U.S. has comparative advantage in wheat and Brazil has comparative advantage in sugar cane.

When a country can produce a good at a lower opportunity cost than another country, we say that this country has a **comparative advantage** in that good. In our example, Brazil has a comparative advantage in sugar cane and the U.S. has a comparative advantage in wheat. One can easily see this with a simple observation of the extreme production points in the PPFs of the two countries. If Brazil devoted all of its resources to producing wheat, it would be producing at point A. If however it had devoted all of its resources to producing sugar cane instead, it would be producing a much larger amount, at point B. By moving from point A to point B Brazil would give up a relatively small quantity in wheat production to obtain a large production in sugar cane. The opposite is true for the U.S. If the U.S. moved

from point A to B and produced only sugar cane, this would result in a large opportunity cost in terms of foregone wheat production.

The slope of the PPF gives the opportunity cost of producing an additional unit of wheat. While the slope is not constant throughout the PPFs, it is quite apparent that the PPF in Brazil is much steeper than in the U.S., and therefore the opportunity cost of wheat generally higher in Brazil. In the chapter on **International Trade** you will learn that countries' differences in comparative advantage determine which goods they will choose to produce and trade. When countries engage in trade, they specialize in the production of the goods in which they have comparative advantage, and trade part of that production for goods in which they do not have comparative advantage. With trade, manufacturers produce goods where the opportunity cost is lowest, so total production increases, benefiting both trading parties.

2.3 | Confronting Objections to the Economic Approach

By the end of this section, you will be able to:

- Analyze arguments against economic approaches to decision-making
- Interpret a tradeoff diagram
- Contrast normative statements and positive statements

It is one thing to understand the economic approach to decision-making and another thing to feel comfortable applying it. The sources of discomfort typically fall into two categories: that people do not act in the way that fits the economic way of thinking, and that even if people did act that way, they should try not to. Let's consider these arguments in turn.

First Objection: People, Firms, and Society Do Not Act Like This

The economic approach to decision-making seems to require more information than most individuals possess and more careful decision-making than most individuals actually display. After all, do you or any of your friends draw a budget constraint and mutter to yourself about maximizing utility before you head to the shopping mall? Do members of the U.S. Congress contemplate production possibilities frontiers before they vote on the annual budget? The messy ways in which people and societies operate somehow doesn't look much like neat budget constraints or smoothly curving production possibilities frontiers.

However, the economics approach can be a useful way to analyze and understand the tradeoffs of economic decisions. To appreciate this point, imagine for a moment that you are playing basketball, dribbling to the right, and throwing a bounce-pass to the left to a teammate who is running toward the basket. A physicist or engineer could work out the correct speed and trajectory for the pass, given the different movements involved and the weight and bounciness of the ball. However, when you are playing basketball, you do not perform any of these calculations. You just pass the ball, and if you are a good player, you will do so with high accuracy.

Someone might argue: "The scientist's formula of the bounce-pass requires a far greater knowledge of physics and far more specific information about speeds of movement and weights than the basketball player actually has, so it must be an unrealistic description of how basketball passes actually occur." This reaction would be wrongheaded. The fact that a good player can throw the ball accurately because of practice and skill, without making a physics calculation, does not mean that the physics calculation is wrong.

Similarly, from an economic point of view, someone who shops for groceries every week has a great deal of practice with how to purchase the combination of goods that will provide that person with utility, even if the shopper does not phrase decisions in terms of a budget constraint. Government institutions may work imperfectly and slowly, but in general, a democratic form of government feels pressure from voters and social institutions to make the choices that are most widely preferred by people in that society. Thus, when thinking about the economic actions of groups of people, firms, and society, it is reasonable, as a first approximation, to analyze them with the tools of economic analysis. For more on this, read about behavioral economics in the chapter on **Consumer Choices**.

Second Objection: People, Firms, and Society Should Not Act This Way

The economics approach portrays people as self-interested. For some critics of this approach, even if self-interest is an accurate description of how people behave, these behaviors are not moral. Instead, the critics argue that people

should be taught to care more deeply about others. Economists offer several answers to these concerns.

First, economics is not a form of moral instruction. Rather, it seeks to describe economic behavior as it actually exists. Philosophers draw a distinction between **positive statements**, which describe the world as it is, and **normative statements**, which describe how the world should be. Positive statements are factual. They may be true or false, but we can test them, at least in principle. Normative statements are subjective questions of opinion. We cannot test them since we cannot prove opinions to be true or false. They just are opinions based on one's values. For example, an economist could analyze a proposed subway system in a certain city. If the expected benefits exceed the costs, he concludes that the project is worthy—an example of positive analysis. Another economist argues for extended unemployment compensation during the Great Depression because a rich country like the United States should take care of its less fortunate citizens—an example of normative analysis.

Even if the line between positive and normative statements is not always crystal clear, economic analysis does try to remain rooted in the study of the actual people who inhabit the actual economy. Fortunately however, the assumption that individuals are purely self-interested is a simplification about human nature. In fact, we need to look no further than to Adam Smith, the very father of modern economics to find evidence of this. The opening sentence of his book, *The Theory of Moral Sentiments*, puts it very clearly: "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it." Clearly, individuals are both self-interested and altruistic.

Second, we can label self-interested behavior and profit-seeking with other names, such as personal choice and freedom. The ability to make personal choices about buying, working, and saving is an important personal freedom. Some people may choose high-pressure, high-paying jobs so that they can earn and spend considerable amounts of money on themselves. Others may allocate large portions of their earnings to charity or spend it on their friends and family. Others may devote themselves to a career that can require much time, energy, and expertise but does not offer high financial rewards, like being an elementary school teacher or a social worker. Still others may choose a job that does consume much of their time or provide a high level of income, but still leaves time for family, friends, and contemplation. Some people may prefer to work for a large company; others might want to start their own business. People's freedom to make their own economic choices has a moral value worth respecting.

Clear It Up

Is a diagram by any other name the same?

When you study economics, you may feel buried under an avalanche of diagrams. Your goal should be to recognize the common underlying logic and pattern of the diagrams, not to memorize each one.

This chapter uses only one basic diagram, although we present it with different sets of labels. The consumption budget constraint and the production possibilities frontier for society, as a whole, are the same basic diagram. **Figure 2.6** shows an individual budget constraint and a production possibilities frontier for two goods, Good 1 and Good 2. The tradeoff diagram always illustrates three basic themes: scarcity, tradeoffs, and economic efficiency.

The first theme is scarcity. It is not feasible to have unlimited amounts of both goods. Even if the budget constraint or a PPF shifts, scarcity remains—just at a different level. The second theme is tradeoffs. As depicted in the budget constraint or the production possibilities frontier, it is necessary to forgo some of one good to gain more of the other good. The details of this tradeoff vary. In a budget constraint we determine, the tradeoff is determined by the relative prices of the goods: that is, the relative price of two goods in the consumption choice budget constraint. These tradeoffs appear as a straight line. However, a curved line represents the tradeoffs in many production possibilities frontiers because the law of diminishing returns holds that as we add resources to an area, the marginal gains tend to diminish. Regardless of the specific shape, tradeoffs remain.

The third theme is economic efficiency, or getting the most benefit from scarce resources. All choices on the production possibilities frontier show productive efficiency because in such cases, there is no way to increase the quantity of one good without decreasing the quantity of the other. Similarly, when an individual makes a

choice along a budget constraint, there is no way to increase the quantity of one good without decreasing the quantity of the other. The choice on a production possibilities set that is socially preferred, or the choice on an individual's budget constraint that is personally preferred, will display allocative efficiency.

The basic budget constraint/production possibilities frontier diagram will recur throughout this book. Some examples include using these tradeoff diagrams to analyze trade, environmental protection and economic output, equality of incomes and economic output, and the macroeconomic tradeoff between consumption and investment. Do not allow the different labels to confuse you. The budget constraint/production possibilities frontier diagram is always just a tool for thinking carefully about scarcity, tradeoffs, and efficiency in a particular situation.

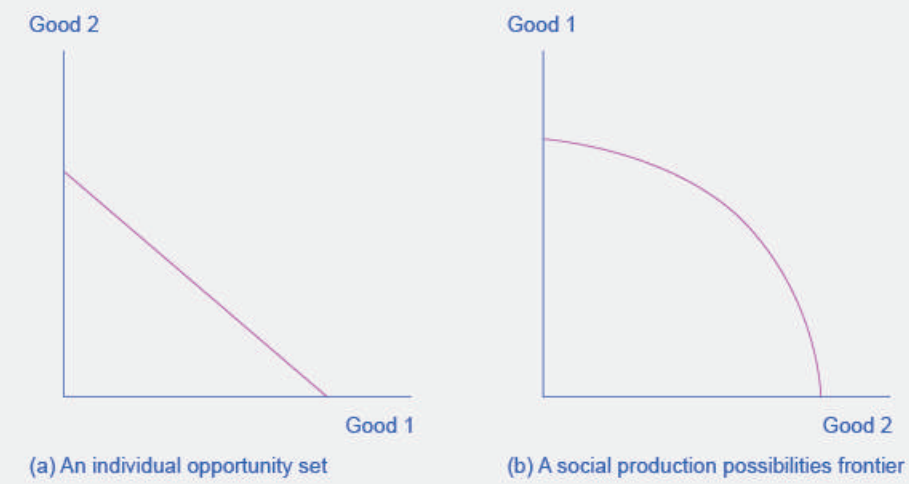


Figure 2.6 The Tradeoff Diagram Both the individual opportunity set (or budget constraint) and the social production possibilities frontier show the constraints under which individual consumers and society as a whole operate. Both diagrams show the tradeoff in choosing more of one good at the cost of less of the other.

Third, self-interested behavior can lead to positive social results. For example, when people work hard to make a living, they create economic output. Consumers who are looking for the best deals will encourage businesses to offer goods and services that meet their needs. Adam Smith, writing in *The Wealth of Nations*, named this property the **invisible hand**. In describing how consumers and producers interact in a market economy, Smith wrote:

> Every individual…generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain. And he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention…By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.

The metaphor of the invisible hand suggests the remarkable possibility that broader social good can emerge from selfish individual actions.

Fourth, even people who focus on their own self-interest in the economic part of their life often set aside their own narrow self-interest in other parts of life. For example, you might focus on your own self-interest when asking your employer for a raise or negotiating to buy a car. Then you might turn around and focus on other people when you volunteer to read stories at the local library, help a friend move to a new apartment, or donate money to a charity. Self-interest is a reasonable starting point for analyzing many economic decisions, without needing to imply that people never do anything that is not in their own immediate self-interest.

Bring it Home

Choices ... To What Degree?

What have we learned? We know that scarcity impacts all the choices we make. An economist might argue that people do not obtain a bachelor's or master's degree because they do not have the resources to make those choices or because their incomes are too low and/or the price of these degrees is too high. A bachelor's or a master's degree may not be available in their opportunity set.

The price of these degrees may be too high not only because the actual price, college tuition (and perhaps room and board), is too high. An economist might also say that for many people, the full opportunity cost of a bachelor's or a master's degree is too high. For these people, they are unwilling or unable to make the tradeoff of forfeiting years of working, and earning an income, to earn a degree.

Finally, the statistics we introduced at the start of the chapter reveal information about intertemporal choices. An economist might say that people choose not to obtain a college degree because they may have to borrow money to attend college, and the interest they have to pay on that loan in the future will affect their decisions today. Also, it could be that some people have a preference for current consumption over future consumption, so they choose to work now at a lower salary and consume now, rather than postponing that consumption until after they graduate college.

KEY TERMS

allocative efficiency when the mix of goods produced represents the mix that society most desires

budget constraint all possible consumption combinations of goods that someone can afford, given the prices of goods, when all income is spent; the boundary of the opportunity set

comparative advantage when a country can produce a good at a lower cost in terms of other goods; or, when a country has a lower opportunity cost of production

invisible hand Adam Smith's concept that individuals' self-interested behavior can lead to positive social outcomes

law of diminishing marginal utility as we consume more of a good or service, the utility we get from additional units of the good or service tends to become smaller than what we received from earlier units

law of diminishing returns as we add additional increments of resources to producing a good or service, the marginal benefit from those additional increments will decline

marginal analysis examination of decisions on the margin, meaning a little more or a little less from the status quo

normative statement statement which describes how the world should be

opportunity cost measures cost by what we give up/forfeit in exchange; opportunity cost measures the value of the forgone alternative

opportunity set all possible combinations of consumption that someone can afford given the prices of goods and the individual’s income

positive statement statement which describes the world as it is

production possibilities frontier (PPF) a diagram that shows the productively efficient combinations of two products that an economy can produce given the resources it has available.

productive efficiency when it is impossible to produce more of one good (or service) without decreasing the quantity produced of another good (or service)

sunk costs costs that we make in the past that we cannot recover

utility satisfaction, usefulness, or value one obtains from consuming goods and services

KEY CONCEPTS AND SUMMARY

2.1 How Individuals Make Choices Based on Their Budget Constraint

Economists see the real world as one of scarcity: that is, a world in which people’s desires exceed what is possible. As a result, economic behavior involves tradeoffs in which individuals, firms, and society must forgo something that they desire to obtain things that they desire more. Individuals face the tradeoff of what quantities of goods and services to consume. The budget constraint, which is the frontier of the opportunity set, illustrates the range of available choices. The relative price of the choices determines the slope of the budget constraint. Choices beyond the budget constraint are not affordable.

Opportunity cost measures cost by what we forgo in exchange. Sometimes we can measure opportunity cost in money, but it is often useful to consider time as well, or to measure it in terms of the actual resources that we must forfeit.

Most economic decisions and tradeoffs are not all-or-nothing. Instead, they involve marginal analysis, which means they are about decisions on the margin, involving a little more or a little less. The law of diminishing marginal utility points out that as a person receives more of something—whether it is a specific good or another resource—the

additional marginal gains tend to become smaller. Because sunk costs occurred in the past and cannot be recovered, they should be disregarded in making current decisions.

2.2 The Production Possibilities Frontier and Social Choices

A production possibilities frontier defines the set of choices society faces for the combinations of goods and services it can produce given the resources available. The shape of the PPF is typically curved outward, rather than straight. Choices outside the PPF are unattainable and choices inside the PPF are wasteful. Over time, a growing economy will tend to shift the PPF outwards.

The law of diminishing returns holds that as increments of additional resources are devoted to producing something, the marginal increase in output will become increasingly smaller. All choices along a production possibilities frontier display productive efficiency; that is, it is impossible to use society's resources to produce more of one good without decreasing production of the other good. The specific choice along a production possibilities frontier that reflects the mix of goods society prefers is the choice with allocative efficiency. The curvature of the PPF is likely to differ by country, which results in different countries having comparative advantage in different goods. Total production can increase if countries specialize in the goods in which they have comparative advantage and trade some of their production for the remaining goods.

2.3 Confronting Objections to the Economic Approach

The economic way of thinking provides a useful approach to understanding human behavior. Economists make the careful distinction between positive statements, which describe the world as it is, and normative statements, which describe how the world should be. Even when economics analyzes the gains and losses from various events or policies, and thus draws normative conclusions about how the world should be, the analysis of economics is rooted in a positive analysis of how people, firms, and governments actually behave, not how they should behave.

SELF-CHECK QUESTIONS

1. Suppose Alphonso's town raised the price of bus tickets to $1 per trip (while the price of burgers stayed at $2 and his budget remained $10 per week.) Draw Alphonso's new budget constraint. What happens to the opportunity cost of bus tickets?

2. Return to the example in **Figure 2.4**. Suppose there is an improvement in medical technology that enables more healthcare with the same amount of resources. How would this affect the production possibilities curve and, in particular, how would it affect the opportunity cost of education?

3. Could a nation be producing in a way that is allocatively efficient, but productively inefficient?

4. What are the similarities between a consumer's budget constraint and society's production possibilities frontier, not just graphically but analytically?

5. Individuals may not act in the rational, calculating way described by the economic model of decision making, measuring utility and costs at the margin, but can you make a case that they behave approximately that way?

6. Would an op-ed piece in a newspaper urging the adoption of a particular economic policy be a positive or normative statement?

7. Would a research study on the effects of soft drink consumption on children's cognitive development be a positive or normative statement?

REVIEW QUESTIONS

8. Explain why scarcity leads to tradeoffs.

9. Explain why individuals make choices that are directly on the budget constraint, rather than inside the budget constraint or outside it.

10. What is comparative advantage?

11. What does a production possibilities frontier illustrate?

12. Why is a production possibilities frontier typically drawn as a curve, rather than a straight line?

13. Explain why societies cannot make a choice above their production possibilities frontier and should not make a choice below it.

14. What are diminishing marginal returns?

15. What is productive efficiency? Allocative efficiency?

16. What is the difference between a positive and a normative statement?

17. Is the economic model of decision-making intended as a literal description of how individuals, firms, and the governments actually make decisions?

18. What are four responses to the claim that people should not behave in the way described in this chapter?

CRITICAL THINKING QUESTIONS

19. Suppose Alphonso's town raises the price of bus tickets from $0.50 to $1 and the price of burgers rises from $2 to $4. Why is the opportunity cost of bus tickets unchanged? Suppose Alphonso's weekly spending money increases from $10 to $20. How is his budget constraint affected from all three changes? Explain.

20. During the Second World War, Germany's factories were decimated. It also suffered many human casualties, both soldiers and civilians. How did the war affect Germany's production possibilities curve?

21. It is clear that productive inefficiency is a waste since resources are used in a way that produces less goods and services than a nation is capable of. Why is allocative inefficiency also wasteful?

22. What assumptions about the economy must be true for the invisible hand to work? To what extent are those assumptions valid in the real world?

23. Do economists have any particular expertise at making normative arguments? In other words, they have expertise at making positive statements (i.e., what *will* happen) about some economic policy, for example, but do they have special expertise to judge whether or not the policy *should* be undertaken?

PROBLEMS

Use this information to answer the following 4 questions: Marie has a weekly budget of $24, which she likes to spend on magazines and pies.

24. If the price of a magazine is $4 each, what is the maximum number of magazines she could buy in a week?

25. If the price of a pie is $12, what is the maximum number of pies she could buy in a week?

26. Draw Marie's budget constraint with pies on the horizontal axis and magazines on the vertical axis. What is the slope of the budget constraint?

27. What is Marie's opportunity cost of purchasing a pie?

3 | Demand and Supply

Figure 3.1 Farmer's Market Organic vegetables and fruits that are grown and sold within a specific geographical region should, in theory, cost less than conventional produce because the transportation costs are less. That is not, however, usually the case. (Credit: Modification of work by Natalie Maynor/Flickr Creative Commons)

Bring it Home

Why Can We Not Get Enough of Organic?

Organic food is increasingly popular, not just in the United States, but worldwide. At one time, consumers had to go to specialty stores or farmers' markets to find organic produce. Now it is available in most grocery stores. In short, organic is part of the mainstream.

Ever wonder why organic food costs more than conventional food? Why, say, does an organic Fuji apple cost $1.99 a pound, while its conventional counterpart costs $1.49 a pound? The same price relationship is true for just about every organic product on the market. If many organic foods are locally grown, would they not take less time to get to market and therefore be cheaper? What are the forces that keep those prices from coming down? Turns out those forces have quite a bit to do with this chapter's topic: demand and supply.

Introduction to Demand and Supply

In this chapter, you will learn about:

- Demand, Supply, and Equilibrium in Markets for Goods and Services
- Shifts in Demand and Supply for Goods and Services
- Changes in Equilibrium Price and Quantity: The Four-Step Process

- Price Ceilings and Price Floors

An auction bidder pays thousands of dollars for a dress Whitney Houston wore. A collector spends a small fortune for a few drawings by John Lennon. People usually react to purchases like these in two ways: their jaw drops because they think these are high prices to pay for such goods or they think these are rare, desirable items and the amount paid seems right.

Link It Up

Visit this website (http://openstaxcollege.org/l/celebauction) to read a list of bizarre items that have been purchased for their ties to celebrities. These examples represent an interesting facet of demand and supply.

When economists talk about prices, they are less interested in making judgments than in gaining a practical understanding of what determines prices and why prices change. Consider a price most of us contend with weekly: that of a gallon of gas. Why was the average price of gasoline in the United States $3.71 per gallon in June 2014? Why did the price for gasoline fall sharply to $1.96 per gallon by January 2016? To explain these price movements, economists focus on the determinants of what gasoline buyers are willing to pay and what gasoline sellers are willing to accept.

As it turns out, the price of gasoline in June of any given year is nearly always higher than the price in January of that same year. Over recent decades, gasoline prices in midsummer have averaged about 10 cents per gallon more than their midwinter low. The likely reason is that people drive more in the summer, and are also willing to pay more for gas, but that does not explain how steeply gas prices fell. Other factors were at work during those six months, such as increases in supply and decreases in the demand for crude oil.

This chapter introduces the economic model of demand and supply—one of the most powerful models in all of economics. The discussion here begins by examining how demand and supply determine the price and the quantity sold in markets for goods and services, and how changes in demand and supply lead to changes in prices and quantities.

3.1 | Demand, Supply, and Equilibrium in Markets for Goods and Services

By the end of this section, you will be able to:

- Explain demand, quantity demanded, and the law of demand
- Identify a demand curve and a supply curve
- Explain supply, quantity supplied, and the law of supply
- Explain equilibrium, equilibrium price, and equilibrium quantity

First let’s first focus on what economists mean by demand, what they mean by supply, and then how demand and supply interact in a market.

Demand for Goods and Services

Economists use the term **demand** to refer to the amount of some good or service consumers are willing and able to purchase at each price. Demand is fundamentally based on needs and wants—if you have no need or want for something, you won't buy it. While a consumer may be able to differentiate between a need and a want, but from an economist's perspective they are the same thing. Demand is also based on ability to pay. If you cannot pay for it, you have no effective demand. By this definition, a homeless person probably has no effective demand for shelter.

What a buyer pays for a unit of the specific good or service is called **price**. The total number of units that consumers would purchase at that price is called the **quantity demanded**. A rise in price of a good or service almost always decreases the quantity demanded of that good or service. Conversely, a fall in price will increase the quantity demanded. When the price of a gallon of gasoline increases, for example, people look for ways to reduce their consumption by combining several errands, commuting by carpool or mass transit, or taking weekend or vacation trips closer to home. Economists call this inverse relationship between price and quantity demanded the **law of demand**. The law of demand assumes that all other variables that affect demand (which we explain in the next module) are held constant.

We can show an example from the market for gasoline in a table or a graph. Economist call a table that shows the quantity demanded at each price, such as **Table 3.1**, a **demand schedule**. In this case we measure price in dollars per gallon of gasoline. We measure the quantity demanded in millions of gallons over some time period (for example, per day or per year) and over some geographic area (like a state or a country). A **demand curve** shows the relationship between price and quantity demanded on a graph like **Figure 3.2**, with quantity on the horizontal axis and the price per gallon on the vertical axis. (Note that this is an exception to the normal rule in mathematics that the independent variable (x) goes on the horizontal axis and the dependent variable (y) goes on the vertical. Economics is not math.)

Table 3.1 shows the demand schedule and the graph in **Figure 3.2** shows the demand curve. These are two ways to describe the same relationship between price and quantity demanded.

Price (per gallon)	Quantity Demanded (millions of gallons)
$1.00	800
$1.20	700
$1.40	600
$1.60	550
$1.80	500
$2.00	460
$2.20	420

Table 3.1 Price and Quantity Demanded of Gasoline

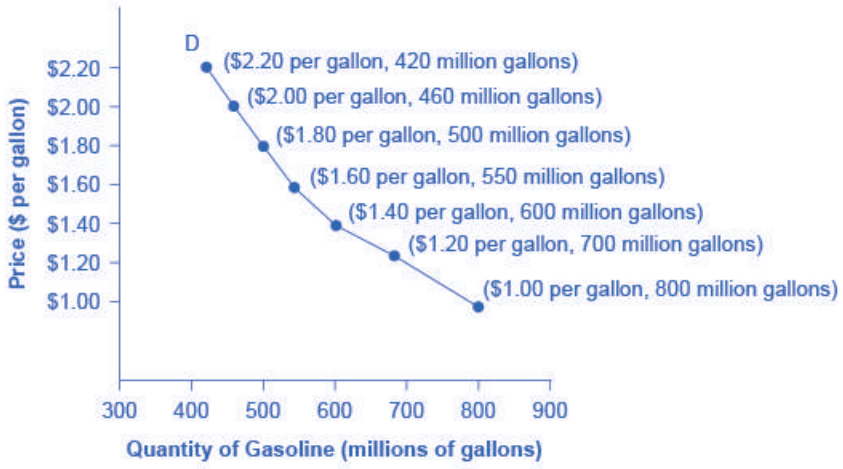


Figure 3.2 A Demand Curve for Gasoline The demand schedule shows that as price rises, quantity demanded decreases, and vice versa. We graph these points, and the line connecting them is the demand curve (D). The downward slope of the demand curve again illustrates the law of demand—the inverse relationship between prices and quantity demanded.

Demand curves will appear somewhat different for each product. They may appear relatively steep or flat, or they may be straight or curved. Nearly all demand curves share the fundamental similarity that they slope down from left to right. Demand curves embody the law of demand: As the price increases, the quantity demanded decreases, and conversely, as the price decreases, the quantity demanded increases.

Confused about these different types of demand? Read the next Clear It Up feature.

Clear It Up

Is demand the same as quantity demanded?

In economic terminology, demand is not the same as quantity demanded. When economists talk about demand, they mean the relationship between a range of prices and the quantities demanded at those prices, as illustrated by a demand curve or a demand schedule. When economists talk about quantity demanded, they mean only a certain point on the demand curve, or one quantity on the demand schedule. In short, demand refers to the curve and quantity demanded refers to the (specific) point on the curve.

Supply of Goods and Services

When economists talk about **supply**, they mean the amount of some good or service a producer is willing to supply at each price. Price is what the producer receives for selling one unit of a good or service. A rise in price almost always leads to an increase in the **quantity supplied** of that good or service, while a fall in price will decrease the quantity supplied. When the price of gasoline rises, for example, it encourages profit-seeking firms to take several actions: expand exploration for oil reserves; drill for more oil; invest in more pipelines and oil tankers to bring the oil to plants for refining into gasoline; build new oil refineries; purchase additional pipelines and trucks to ship the gasoline to gas stations; and open more gas stations or keep existing gas stations open longer hours. Economists call this positive relationship between price and quantity supplied—that a higher price leads to a higher quantity supplied and a lower price leads to a lower quantity supplied—the **law of supply**. The law of supply assumes that all other variables that affect supply (to be explained in the next module) are held constant.

Still unsure about the different types of supply? See the following Clear It Up feature.

Clear It Up

Is supply the same as quantity supplied?

In economic terminology, supply is not the same as quantity supplied. When economists refer to supply, they mean the relationship between a range of prices and the quantities supplied at those prices, a relationship that we can illustrate with a supply curve or a supply schedule. When economists refer to quantity supplied, they mean only a certain point on the supply curve, or one quantity on the supply schedule. In short, supply refers to the curve and quantity supplied refers to the (specific) point on the curve.

Figure 3.3 illustrates the law of supply, again using the market for gasoline as an example. Like demand, we can illustrate supply using a table or a graph. A **supply schedule** is a table, like **Table 3.2**, that shows the quantity supplied at a range of different prices. Again, we measure price in dollars per gallon of gasoline and we measure quantity supplied in millions of gallons. A **supply curve** is a graphic illustration of the relationship between price, shown on the vertical axis, and quantity, shown on the horizontal axis. The supply schedule and the supply curve are just two different ways of showing the same information. Notice that the horizontal and vertical axes on the graph for the supply curve are the same as for the demand curve.

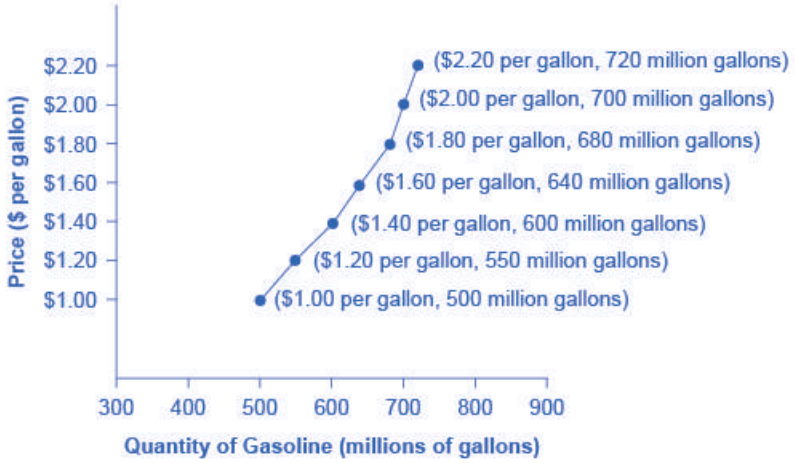


Figure 3.3 A Supply Curve for Gasoline The supply schedule is the table that shows quantity supplied of gasoline at each price. As price rises, quantity supplied also increases, and vice versa. The supply curve (S) is created by graphing the points from the supply schedule and then connecting them. The upward slope of the supply curve illustrates the law of supply—that a higher price leads to a higher quantity supplied, and vice versa.

Price (per gallon)	Quantity Supplied (millions of gallons)
$1.00	500
$1.20	550
$1.40	600
$1.60	640
$1.80	680
$2.00	700

Table 3.2 Price and Supply of Gasoline

Price (per gallon)	Quantity Supplied (millions of gallons)
$2.20	720

Table 3.2 Price and Supply of Gasoline

The shape of supply curves will vary somewhat according to the product: steeper, flatter, straighter, or curved. Nearly all supply curves, however, share a basic similarity: they slope up from left to right and illustrate the law of supply: as the price rises, say, from $1.00 per gallon to $2.20 per gallon, the quantity supplied increases from 500 gallons to 720 gallons. Conversely, as the price falls, the quantity supplied decreases.

Equilibrium—Where Demand and Supply Intersect

Because the graphs for demand and supply curves both have price on the vertical axis and quantity on the horizontal axis, the demand curve and supply curve for a particular good or service can appear on the same graph. Together, demand and supply determine the price and the quantity that will be bought and sold in a market.

Figure 3.4 illustrates the interaction of demand and supply in the market for gasoline. The demand curve (D) is identical to **Figure 3.2**. The supply curve (S) is identical to **Figure 3.3**. **Table 3.3** contains the same information in tabular form.

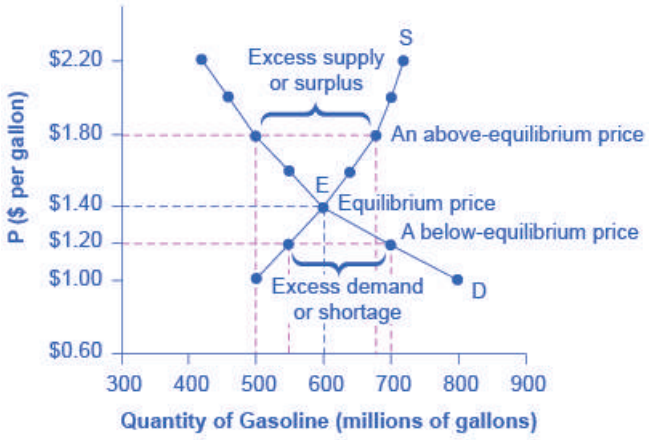


Figure 3.4 Demand and Supply for Gasoline The demand curve (D) and the supply curve (S) intersect at the equilibrium point E, with a price of $1.40 and a quantity of 600. The equilibrium is the only price where quantity demanded is equal to quantity supplied. At a price above equilibrium like $1.80, quantity supplied exceeds the quantity demanded, so there is excess supply. At a price below equilibrium such as $1.20, quantity demanded exceeds quantity supplied, so there is excess demand.

Price (per gallon)	Quantity demanded (millions of gallons)	Quantity supplied (millions of gallons)
$1.00	800	500
$1.20	700	550
$1.40	**600**	**600**
$1.60	550	640
$1.80	500	680

Table 3.3 Price, Quantity Demanded, and Quantity Supplied

Price (per gallon)	Quantity demanded (millions of gallons)	Quantity supplied (millions of gallons)
$2.00	460	700
$2.20	420	720

Table 3.3 Price, Quantity Demanded, and Quantity Supplied

Remember this: When two lines on a diagram cross, this intersection usually means something. The point where the supply curve (S) and the demand curve (D) cross, designated by point E in **Figure 3.4**, is called the **equilibrium**. The **equilibrium price** is the only price where the plans of consumers and the plans of producers agree—that is, where the amount of the product consumers want to buy (quantity demanded) is equal to the amount producers want to sell (quantity supplied). Economists call this common quantity the **equilibrium quantity**. At any other price, the quantity demanded does not equal the quantity supplied, so the market is not in equilibrium at that price.

In **Figure 3.4**, the equilibrium price is $1.40 per gallon of gasoline and the equilibrium quantity is 600 million gallons. If you had only the demand and supply schedules, and not the graph, you could find the equilibrium by looking for the price level on the tables where the quantity demanded and the quantity supplied are equal.

The word “equilibrium” means “balance.” If a market is at its equilibrium price and quantity, then it has no reason to move away from that point. However, if a market is not at equilibrium, then economic pressures arise to move the market toward the equilibrium price and the equilibrium quantity.

Imagine, for example, that the price of a gallon of gasoline was above the equilibrium price—that is, instead of $1.40 per gallon, the price is $1.80 per gallon. The dashed horizontal line at the price of $1.80 in **Figure 3.4** illustrates this above equilibrium price. At this higher price, the quantity demanded drops from 600 to 500. This decline in quantity reflects how consumers react to the higher price by finding ways to use less gasoline.

Moreover, at this higher price of $1.80, the quantity of gasoline supplied rises from the 600 to 680, as the higher price makes it more profitable for gasoline producers to expand their output. Now, consider how quantity demanded and quantity supplied are related at this above-equilibrium price. Quantity demanded has fallen to 500 gallons, while quantity supplied has risen to 680 gallons. In fact, at any above-equilibrium price, the quantity supplied exceeds the quantity demanded. We call this an **excess supply** or a **surplus**.

With a surplus, gasoline accumulates at gas stations, in tanker trucks, in pipelines, and at oil refineries. This accumulation puts pressure on gasoline sellers. If a surplus remains unsold, those firms involved in making and selling gasoline are not receiving enough cash to pay their workers and to cover their expenses. In this situation, some producers and sellers will want to cut prices, because it is better to sell at a lower price than not to sell at all. Once some sellers start cutting prices, others will follow to avoid losing sales. These price reductions in turn will stimulate a higher quantity demanded. Therefore, if the price is above the equilibrium level, incentives built into the structure of demand and supply will create pressures for the price to fall toward the equilibrium.

Now suppose that the price is below its equilibrium level at $1.20 per gallon, as the dashed horizontal line at this price in **Figure 3.4** shows. At this lower price, the quantity demanded increases from 600 to 700 as drivers take longer trips, spend more minutes warming up the car in the driveway in wintertime, stop sharing rides to work, and buy larger cars that get fewer miles to the gallon. However, the below-equilibrium price reduces gasoline producers’ incentives to produce and sell gasoline, and the quantity supplied falls from 600 to 550.

When the price is below equilibrium, there is **excess demand**, or a **shortage**—that is, at the given price the quantity demanded, which has been stimulated by the lower price, now exceeds the quantity supplied, which had been depressed by the lower price. In this situation, eager gasoline buyers mob the gas stations, only to find many stations running short of fuel. Oil companies and gas stations recognize that they have an opportunity to make higher profits by selling what gasoline they have at a higher price. As a result, the price rises toward the equilibrium level. Read **Demand, Supply, and Efficiency** for more discussion on the importance of the demand and supply model.

3.2 | Shifts in Demand and Supply for Goods and Services

By the end of this section, you will be able to:

- Identify factors that affect demand
- Graph demand curves and demand shifts
- Identify factors that affect supply
- Graph supply curves and supply shifts

The previous module explored how price affects the quantity demanded and the quantity supplied. The result was the demand curve and the supply curve. Price, however, is not the only factor that influences demand, nor is it the only thing that influences supply. For example, how is demand for vegetarian food affected if, say, health concerns cause more consumers to avoid eating meat? How is the supply of diamonds affected if diamond producers discover several new diamond mines? What are the major factors, in addition to the price, that influence demand or supply?

Link It Up

Visit this website (http://openstaxcollege.org/l/toothfish) to read a brief note on how marketing strategies can influence supply and demand of products.

What Factors Affect Demand?

We defined demand as the amount of some product a consumer is willing and able to purchase at each price. That suggests at least two factors in addition to price that affect demand. Willingness to purchase suggests a desire, based on what economists call tastes and preferences. If you neither need nor want something, you will not buy it. Ability to purchase suggests that income is important. Professors are usually able to afford better housing and transportation than students, because they have more income. Prices of related goods can affect demand also. If you need a new car, the price of a Honda may affect your demand for a Ford. Finally, the size or composition of the population can affect demand. The more children a family has, the greater their demand for clothing. The more driving-age children a family has, the greater their demand for car insurance, and the less for diapers and baby formula.

These factors matter for both individual and market demand as a whole. Exactly how do these various factors affect demand, and how do we show the effects graphically? To answer those questions, we need the *ceteris paribus* assumption.

The *Ceteris Paribus* Assumption

A demand curve or a supply curve is a relationship between two, and only two, variables: quantity on the horizontal axis and price on the vertical axis. The assumption behind a demand curve or a supply curve is that no relevant economic factors, other than the product's price, are changing. Economists call this assumption **ceteris paribus**, a Latin phrase meaning "other things being equal." Any given demand or supply curve is based on the *ceteris paribus* assumption that all else is held equal. A demand curve or a supply curve is a relationship between two, and only two, variables when all other variables are kept constant. If all else is not held equal, then the laws of supply and demand will not necessarily hold, as the following Clear It Up feature shows.

Clear It Up

When does *ceteris paribus* apply?

We typically apply *ceteris paribus* when we observe how changes in price affect demand or supply, but we can apply *ceteris paribus* more generally. In the real world, demand and supply depend on more factors than just price. For example, a consumer's demand depends on income and a producer's supply depends on the cost of producing the product. How can we analyze the effect on demand or supply if multiple factors are changing at the same time—say price rises and income falls? The answer is that we examine the changes one at a time, assuming the other factors are held constant.

For example, we can say that an increase in the price reduces the amount consumers will buy (assuming income, and anything else that affects demand, is unchanged). Additionally, a decrease in income reduces the amount consumers can afford to buy (assuming price, and anything else that affects demand, is unchanged). This is what the *ceteris paribus* assumption really means. In this particular case, after we analyze each factor separately, we can combine the results. The amount consumers buy falls for two reasons: first because of the higher price and second because of the lower income.

How Does Income Affect Demand?

Let's use income as an example of how factors other than price affect demand. **Figure 3.5** shows the initial demand for automobiles as D_0. At point Q, for example, if the price is $20,000 per car, the quantity of cars demanded is 18 million. D_0 also shows how the quantity of cars demanded would change as a result of a higher or lower price. For example, if the price of a car rose to $22,000, the quantity demanded would decrease to 17 million, at point R.

The original demand curve D_0, like every demand curve, is based on the *ceteris paribus* assumption that no other economically relevant factors change. Now imagine that the economy expands in a way that raises the incomes of many people, making cars more affordable. How will this affect demand? How can we show this graphically?

Return to **Figure 3.5**. The price of cars is still $20,000, but with higher incomes, the quantity demanded has now increased to 20 million cars, shown at point S. As a result of the higher income levels, the demand curve shifts to the right to the new demand curve D_1, indicating an increase in demand. **Table 3.4** shows clearly that this increased demand would occur at every price, not just the original one.

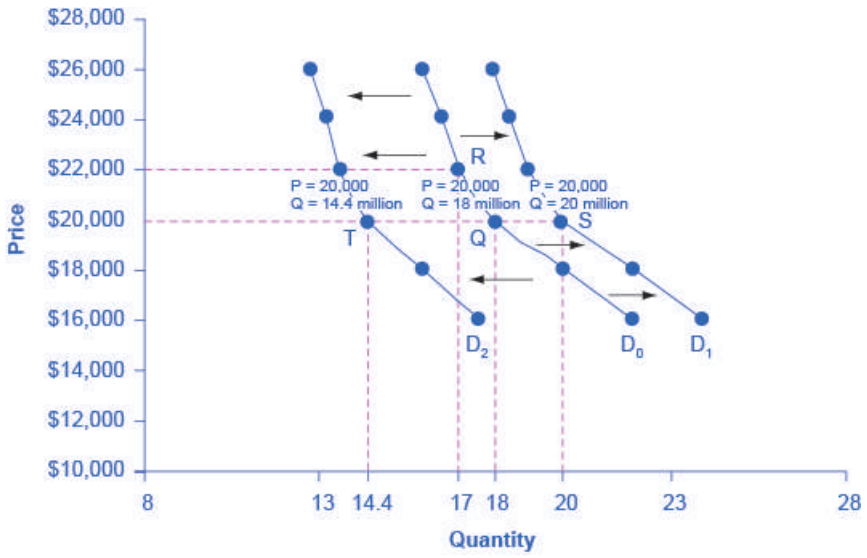


Figure 3.5 Shifts in Demand: A Car Example Increased demand means that at every given price, the quantity demanded is higher, so that the demand curve shifts to the right from D_0 to D_1. Decreased demand means that at every given price, the quantity demanded is lower, so that the demand curve shifts to the left from D_0 to D_2.

Price	Decrease to D_2	Original Quantity Demanded D_0	Increase to D_1
$16,000	17.6 million	22.0 million	24.0 million
$18,000	16.0 million	20.0 million	22.0 million
$20,000	14.4 million	18.0 million	20.0 million
$22,000	13.6 million	17.0 million	19.0 million
$24,000	13.2 million	16.5 million	18.5 million
$26,000	12.8 million	16.0 million	18.0 million

Table 3.4 Price and Demand Shifts: A Car Example

Now, imagine that the economy slows down so that many people lose their jobs or work fewer hours, reducing their incomes. In this case, the decrease in income would lead to a lower quantity of cars demanded at every given price, and the original demand curve D_0 would shift left to D_2. The shift from D_0 to D_2 represents such a decrease in demand: At any given price level, the quantity demanded is now lower. In this example, a price of $20,000 means 18 million cars sold along the original demand curve, but only 14.4 million sold after demand fell.

When a demand curve shifts, it does not mean that the quantity demanded by every individual buyer changes by the same amount. In this example, not everyone would have higher or lower income and not everyone would buy or not buy an additional car. Instead, a shift in a demand curve captures a pattern for the market as a whole.

In the previous section, we argued that higher income causes greater demand at every price. This is true for most goods and services. For some—luxury cars, vacations in Europe, and fine jewelry—the effect of a rise in income can be especially pronounced. A product whose demand rises when income rises, and vice versa, is called a **normal good**. A few exceptions to this pattern do exist. As incomes rise, many people will buy fewer generic brand groceries and more name brand groceries. They are less likely to buy used cars and more likely to buy new cars. They will be less likely to rent an apartment and more likely to own a home. A product whose demand falls when income rises, and vice versa, is called an **inferior good**. In other words, when income increases, the demand curve shifts to the left.

Other Factors That Shift Demand Curves

Income is not the only factor that causes a shift in demand. Other factors that change demand include tastes and preferences, the composition or size of the population, the prices of related goods, and even expectations. A change in any one of the underlying factors that determine what quantity people are willing to buy at a given price will cause a shift in demand. Graphically, the new demand curve lies either to the right (an increase) or to the left (a decrease) of the original demand curve. Let’s look at these factors.

Changing Tastes or Preferences

From 1980 to 2014, the per-person consumption of chicken by Americans rose from 48 pounds per year to 85 pounds per year, and consumption of beef fell from 77 pounds per year to 54 pounds per year, according to the U.S. Department of Agriculture (USDA). Changes like these are largely due to movements in taste, which change the quantity of a good demanded at every price: that is, they shift the demand curve for that good, rightward for chicken and leftward for beef.

Changes in the Composition of the Population

The proportion of elderly citizens in the United States population is rising. It rose from 9.8% in 1970 to 12.6% in 2000, and will be a projected (by the U.S. Census Bureau) 20% of the population by 2030. A society with relatively more children, like the United States in the 1960s, will have greater demand for goods and services like tricycles and day care facilities. A society with relatively more elderly persons, as the United States is projected to have by 2030, has a higher demand for nursing homes and hearing aids. Similarly, changes in the size of the population can affect the demand for housing and many other goods. Each of these changes in demand will be shown as a shift in the demand curve.

Changes in the prices of related goods such as substitutes or complements also can affect the demand for a product. A

substitute is a good or service that we can use in place of another good or service. As electronic books, like this one, become more available, you would expect to see a decrease in demand for traditional printed books. A lower price for a substitute decreases demand for the other product. For example, in recent years as the price of tablet computers has fallen, the quantity demanded has increased (because of the law of demand). Since people are purchasing tablets, there has been a decrease in demand for laptops, which we can show graphically as a leftward shift in the demand curve for laptops. A higher price for a substitute good has the reverse effect.

Other goods are **complements** for each other, meaning we often use the goods together, because consumption of one good tends to enhance consumption of the other. Examples include breakfast cereal and milk; notebooks and pens or pencils, golf balls and golf clubs; gasoline and sport utility vehicles; and the five-way combination of bacon, lettuce, tomato, mayonnaise, and bread. If the price of golf clubs rises, since the quantity demanded of golf clubs falls (because of the law of demand), demand for a complement good like golf balls decreases, too. Similarly, a higher price for skis would shift the demand curve for a complement good like ski resort trips to the left, while a lower price for a complement has the reverse effect.

Changes in Expectations about Future Prices or Other Factors that Affect Demand

While it is clear that the price of a good affects the quantity demanded, it is also true that expectations about the future price (or expectations about tastes and preferences, income, and so on) can affect demand. For example, if people hear that a hurricane is coming, they may rush to the store to buy flashlight batteries and bottled water. If people learn that the price of a good like coffee is likely to rise in the future, they may head for the store to stock up on coffee now. We show these changes in demand as shifts in the curve. Therefore, a **shift in demand** happens when a change in some economic factor (other than price) causes a different quantity to be demanded at every price. The following Work It Out feature shows how this happens.

Work It Out

Shift in Demand

A shift in demand means that at any price (and at every price), the quantity demanded will be different than it was before. Following is an example of a shift in demand due to an income increase.

Step 1. Draw the graph of a demand curve for a normal good like pizza. Pick a price (like P_0). Identify the corresponding Q_0. See an example in **Figure 3.6**.

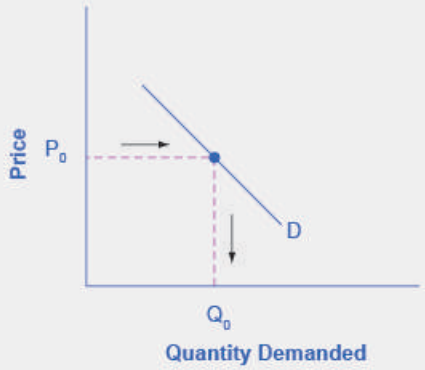


Figure 3.6 Demand Curve We can use the demand curve to identify how much consumers would buy at any given price.

Step 2. Suppose income increases. As a result of the change, are consumers going to buy more or less pizza? The answer is more. Draw a dotted horizontal line from the chosen price, through the original quantity demanded, to the new point with the new Q_1. Draw a dotted vertical line down to the horizontal axis and label the new Q_1. **Figure 3.7** provides an example.

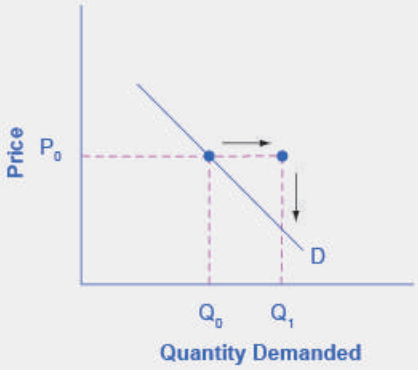


Figure 3.7 Demand Curve with Income Increase With an increase in income, consumers will purchase larger quantities, pushing demand to the right.

Step 3. Now, shift the curve through the new point. You will see that an increase in income causes an upward (or rightward) shift in the demand curve, so that at any price the quantities demanded will be higher, as **Figure 3.8** illustrates.

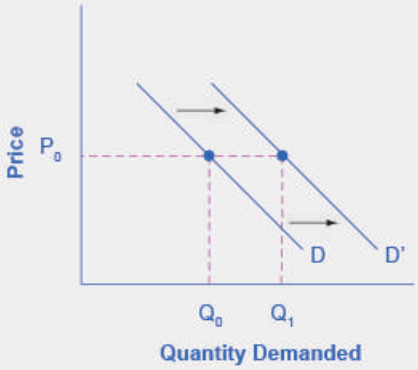


Figure 3.8 Demand Curve Shifted Right With an increase in income, consumers will purchase larger quantities, pushing demand to the right, and causing the demand curve to shift right.

Summing Up Factors That Change Demand

Figure 3.9 summarizes six factors that can shift demand curves. The direction of the arrows indicates whether the demand curve shifts represent an increase in demand or a decrease in demand. Notice that a change in the price of the good or service itself is not listed among the factors that can shift a demand curve. A change in the price of a good or service causes a movement along a specific demand curve, and it typically leads to some change in the quantity demanded, but it does not shift the demand curve.

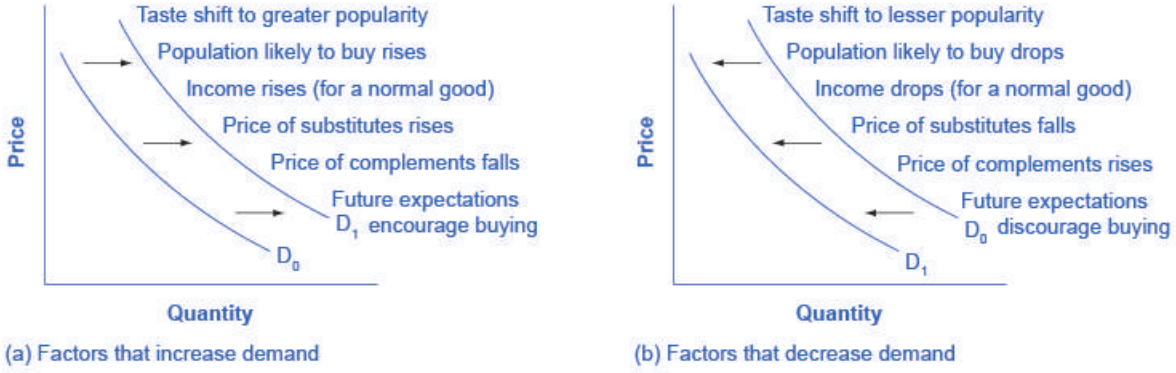


Figure 3.9 Factors That Shift Demand Curves (a) A list of factors that can cause an increase in demand from D_0 to D_1. (b) The same factors, if their direction is reversed, can cause a decrease in demand from D_0 to D_1.

When a demand curve shifts, it will then intersect with a given supply curve at a different equilibrium price and quantity. We are, however, getting ahead of our story. Before discussing how changes in demand can affect equilibrium price and quantity, we first need to discuss shifts in supply curves.

How Production Costs Affect Supply

A supply curve shows how quantity supplied will change as the price rises and falls, assuming *ceteris paribus* so that no other economically relevant factors are changing. If other factors relevant to supply do change, then the entire supply curve will shift. Just as we described a shift in demand as a change in the quantity demanded at every price, a **shift in supply** means a change in the quantity supplied at every price.

In thinking about the factors that affect supply, remember what motivates firms: profits, which are the difference between revenues and costs. A firm produces goods and services using combinations of labor, materials, and machinery, or what we call **inputs** or **factors of production**. If a firm faces lower costs of production, while the prices for the good or service the firm produces remain unchanged, a firm's profits go up. When a firm's profits increase, it is more motivated to produce output, since the more it produces the more profit it will earn. When costs of production fall, a firm will tend to supply a larger quantity at any given price for its output. We can show this by the supply curve shifting to the right.

Take, for example, a messenger company that delivers packages around a city. The company may find that buying gasoline is one of its main costs. If the price of gasoline falls, then the company will find it can deliver messages more cheaply than before. Since lower costs correspond to higher profits, the messenger company may now supply more of its services at any given price. For example, given the lower gasoline prices, the company can now serve a greater area, and increase its supply.

Conversely, if a firm faces higher costs of production, then it will earn lower profits at any given selling price for its products. As a result, a higher cost of production typically causes a firm to supply a smaller quantity at any given price. In this case, the supply curve shifts to the left.

Consider the supply for cars, shown by curve S_0 in **Figure 3.10**. Point J indicates that if the price is $20,000, the quantity supplied will be 18 million cars. If the price rises to $22,000 per car, *ceteris paribus*, the quantity supplied will rise to 20 million cars, as point K on the S_0 curve shows. We can show the same information in table form, as in **Table 3.5**.

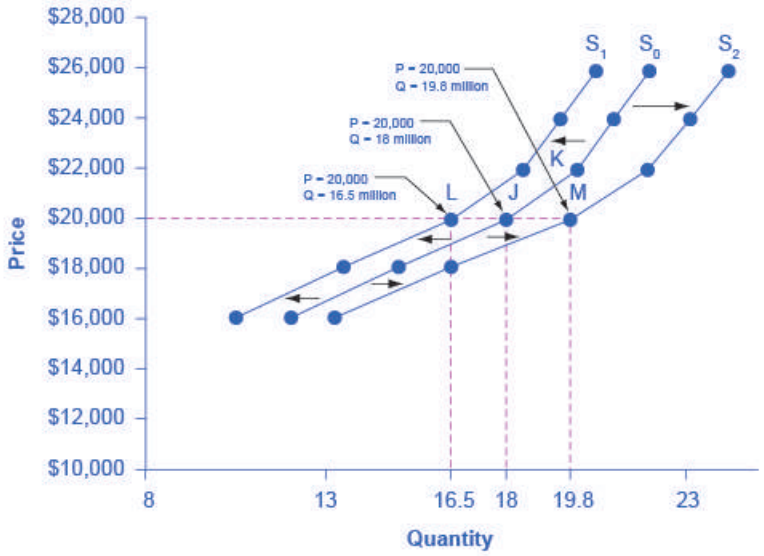


Figure 3.10 Shifts in Supply: A Car Example Decreased supply means that at every given price, the quantity supplied is lower, so that the supply curve shifts to the left, from S_0 to S_1. Increased supply means that at every given price, the quantity supplied is higher, so that the supply curve shifts to the right, from S_0 to S_2.

Price	Decrease to S_1	Original Quantity Supplied S_0	Increase to S_2
$16,000	10.5 million	12.0 million	13.2 million
$18,000	13.5 million	15.0 million	16.5 million
$20,000	16.5 million	18.0 million	19.8 million
$22,000	18.5 million	20.0 million	22.0 million
$24,000	19.5 million	21.0 million	23.1 million
$26,000	20.5 million	22.0 million	24.2 million

Table 3.5 Price and Shifts in Supply: A Car Example

Now, imagine that the price of steel, an important ingredient in manufacturing cars, rises, so that producing a car has become more expensive. At any given price for selling cars, car manufacturers will react by supplying a lower quantity. We can show this graphically as a leftward shift of supply, from S_0 to S_1, which indicates that at any given price, the quantity supplied decreases. In this example, at a price of $20,000, the quantity supplied decreases from 18 million on the original supply curve (S_0) to 16.5 million on the supply curve S_1, which is labeled as point L.

Conversely, if the price of steel decreases, producing a car becomes less expensive. At any given price for selling cars, car manufacturers can now expect to earn higher profits, so they will supply a higher quantity. The shift of supply to the right, from S_0 to S_2, means that at all prices, the quantity supplied has increased. In this example, at a price of $20,000, the quantity supplied increases from 18 million on the original supply curve (S_0) to 19.8 million on the supply curve S_2, which is labeled M.

Other Factors That Affect Supply

In the example above, we saw that changes in the prices of inputs in the production process will affect the cost of production and thus the supply. Several other things affect the cost of production, too, such as changes in weather or other natural conditions, new technologies for production, and some government policies.

Changes in weather and climate will affect the cost of production for many agricultural products. For example, in

2014 the Manchurian Plain in Northeastern China, which produces most of the country's wheat, corn, and soybeans, experienced its most severe drought in 50 years. A drought decreases the supply of agricultural products, which means that at any given price, a lower quantity will be supplied. Conversely, especially good weather would shift the supply curve to the right.

When a firm discovers a new technology that allows the firm to produce at a lower cost, the supply curve will shift to the right, as well. For instance, in the 1960s a major scientific effort nicknamed the Green Revolution focused on breeding improved seeds for basic crops like wheat and rice. By the early 1990s, more than two-thirds of the wheat and rice in low-income countries around the world used these Green Revolution seeds—and the harvest was twice as high per acre. A technological improvement that reduces costs of production will shift supply to the right, so that a greater quantity will be produced at any given price.

Government policies can affect the cost of production and the supply curve through taxes, regulations, and subsidies. For example, the U.S. government imposes a tax on alcoholic beverages that collects about $8 billion per year from producers. Businesses treat taxes as costs. Higher costs decrease supply for the reasons we discussed above. Other examples of policy that can affect cost are the wide array of government regulations that require firms to spend money to provide a cleaner environment or a safer workplace. Complying with regulations increases costs.

A government subsidy, on the other hand, is the opposite of a tax. A subsidy occurs when the government pays a firm directly or reduces the firm's taxes if the firm carries out certain actions. From the firm's perspective, taxes or regulations are an additional cost of production that shifts supply to the left, leading the firm to produce a lower quantity at every given price. Government subsidies reduce the cost of production and increase supply at every given price, shifting supply to the right. The following Work It Out feature shows how this shift happens.

Work It Out

Shift in Supply

We know that a supply curve shows the minimum price a firm will accept to produce a given quantity of output. What happens to the supply curve when the cost of production goes up? Following is an example of a shift in supply due to a production cost increase.

Step 1. Draw a graph of a supply curve for pizza. Pick a quantity (like Q_0). If you draw a vertical line up from Q_0 to the supply curve, you will see the price the firm chooses. **Figure 3.11** provides an example.

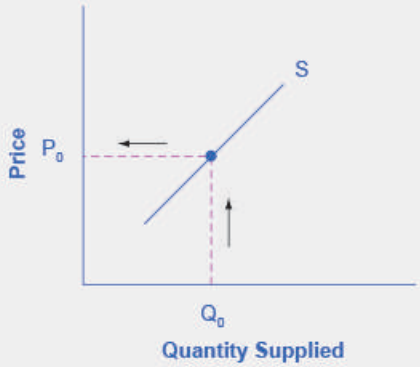


Figure 3.11 Supply Curve You can use a supply curve to show the minimum price a firm will accept to produce a given quantity of output.

Step 2. Why did the firm choose that price and not some other? One way to think about this is that the price is composed of two parts. The first part is the cost of producing pizzas at the margin; in this case, the cost of producing the pizza, including cost of ingredients (e.g., dough, sauce, cheese, and pepperoni), the cost of the pizza oven, the shop rent, and the workers' wages. The second part is the firm's desired profit, which is determined, among other factors, by the profit margins in that particular business. If you add these two parts together, you get the price the firm wishes to charge. The quantity Q0 and associated price P0 give you one point on the firm's supply curve, as **Figure 3.12** illustrates.

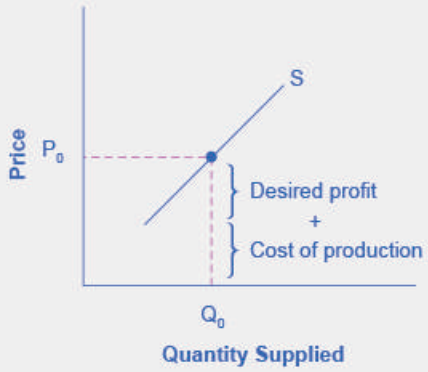


Figure 3.12 Setting Prices The cost of production and the desired profit equal the price a firm will set for a product.

Step 3. Now, suppose that the cost of production increases. Perhaps cheese has become more expensive by $0.75 per pizza. If that is true, the firm will want to raise its price by the amount of the increase in cost ($0.75). Draw this point on the supply curve directly above the initial point on the curve, but $0.75 higher, as **Figure 3.13** shows.

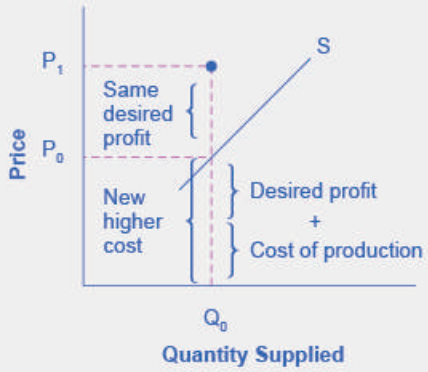


Figure 3.13 Increasing Costs Leads to Increasing Price Because the cost of production and the desired profit equal the price a firm will set for a product, if the cost of production increases, the price for the product will also need to increase.

Step 4. Shift the supply curve through this point. You will see that an increase in cost causes an upward (or a leftward) shift of the supply curve so that at any price, the quantities supplied will be smaller, as **Figure 3.14** illustrates.

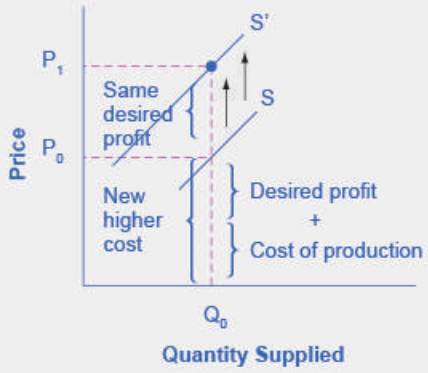


Figure 3.14 Supply Curve Shifts When the cost of production increases, the supply curve shifts upwardly to a new price level.

Summing Up Factors That Change Supply

Changes in the cost of inputs, natural disasters, new technologies, and the impact of government decisions all affect the cost of production. In turn, these factors affect how much firms are willing to supply at any given price.

Figure 3.15 summarizes factors that change the supply of goods and services. Notice that a change in the price of the product itself is not among the factors that shift the supply curve. Although a change in price of a good or service typically causes a change in quantity supplied or a movement along the supply curve for that specific good or service, it does not cause the supply curve itself to shift.

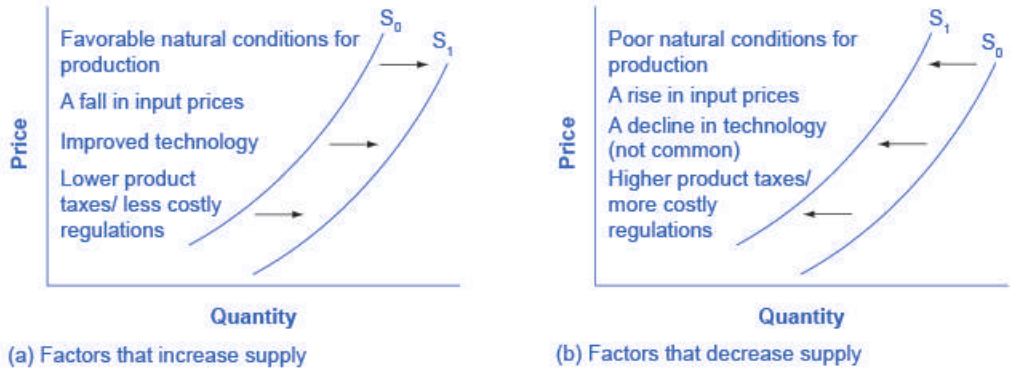


Figure 3.15 Factors That Shift Supply Curves (a) A list of factors that can cause an increase in supply from S_0 to S_1. (b) The same factors, if their direction is reversed, can cause a decrease in supply from S_0 to S_1.

Because demand and supply curves appear on a two-dimensional diagram with only price and quantity on the axes, an unwary visitor to the land of economics might be fooled into believing that economics is about only four topics: demand, supply, price, and quantity. However, demand and supply are really "umbrella" concepts: demand covers all the factors that affect demand, and supply covers all the factors that affect supply. We include factors other than price that affect demand and supply are included by using shifts in the demand or the supply curve. In this way, the two-dimensional demand and supply model becomes a powerful tool for analyzing a wide range of economic circumstances.

3.3 | Changes in Equilibrium Price and Quantity: The Four-Step Process

By the end of this section, you will be able to:

- Identify equilibrium price and quantity through the four-step process
- Graph equilibrium price and quantity
- Contrast shifts of demand or supply and movements along a demand or supply curve
- Graph demand and supply curves, including equilibrium price and quantity, based on real-world examples

Let's begin this discussion with a single economic event. It might be an event that affects demand, like a change in income, population, tastes, prices of substitutes or complements, or expectations about future prices. It might be an event that affects supply, like a change in natural conditions, input prices, or technology, or government policies that affect production. How does this economic event affect equilibrium price and quantity? We will analyze this question using a four-step process.

Step 1. Draw a demand and supply model before the economic change took place. To establish the model requires four standard pieces of information: The law of demand, which tells us the slope of the demand curve; the law of supply, which gives us the slope of the supply curve; the shift variables for demand; and the shift variables for supply. From this model, find the initial equilibrium values for price and quantity.

Step 2. Decide whether the economic change you are analyzing affects demand or supply. In other words, does the

event refer to something in the list of demand factors or supply factors?

Step 3. Decide whether the effect on demand or supply causes the curve to shift to the right or to the left, and sketch the new demand or supply curve on the diagram. In other words, does the event increase or decrease the amount consumers want to buy or producers want to sell?

Step 4. Identify the new equilibrium and then compare the original equilibrium price and quantity to the new equilibrium price and quantity.

Let's consider one example that involves a shift in supply and one that involves a shift in demand. Then we will consider an example where both supply and demand shift.

Good Weather for Salmon Fishing

Supposed that during the summer of 2015, weather conditions were excellent for commercial salmon fishing off the California coast. Heavy rains meant higher than normal levels of water in the rivers, which helps the salmon to breed. Slightly cooler ocean temperatures stimulated the growth of plankton, the microscopic organisms at the bottom of the ocean food chain, providing everything in the ocean with a hearty food supply. The ocean stayed calm during fishing season, so commercial fishing operations did not lose many days to bad weather. How did these climate conditions affect the quantity and price of salmon? **Figure 3.16** illustrates the four-step approach, which we explain below, to work through this problem. **Table 3.6** also provides the information to work the problem.

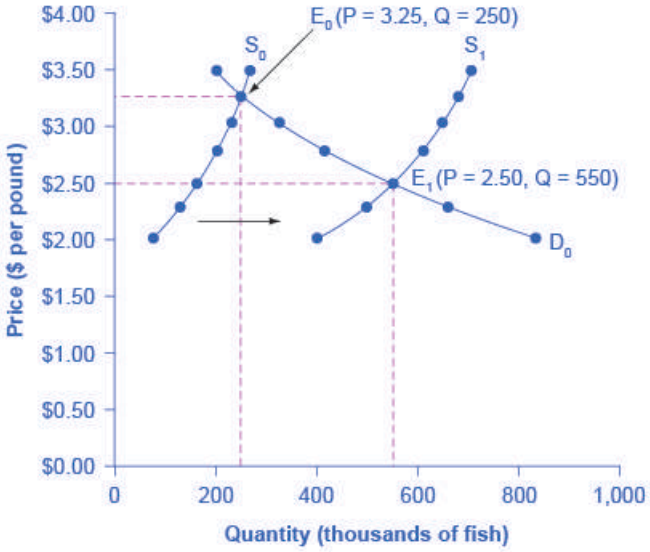


Figure 3.16 Good Weather for Salmon Fishing: The Four-Step Process Unusually good weather leads to changes in the price and quantity of salmon.

Price per Pound	Quantity Supplied in 2014	Quantity Supplied in 2015	Quantity Demanded
$2.00	80	400	840
$2.25	120	480	680
$2.50	160	550	550
$2.75	200	600	450
$3.00	230	640	350
$3.25	250	670	250

Table 3.6 Salmon Fishing

Price per Pound	Quantity Supplied in 2014	Quantity Supplied in 2015	Quantity Demanded
$3.50	270	700	200

Table 3.6 Salmon Fishing

Step 1. Draw a demand and supply model to illustrate the market for salmon in the year before the good weather conditions began. The demand curve D_0 and the supply curve S_0 show that the original equilibrium price is $3.25 per pound and the original equilibrium quantity is 250,000 fish. (This price per pound is what commercial buyers pay at the fishing docks. What consumers pay at the grocery is higher.)

Step 2. Did the economic event affect supply or demand? Good weather is an example of a natural condition that affects supply.

Step 3. Was the effect on supply an increase or a decrease? Good weather is a change in natural conditions that increases the quantity supplied at any given price. The supply curve shifts to the right, moving from the original supply curve S_0 to the new supply curve S_1, which **Figure 3.16** and **Table 3.6** show.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium. At the new equilibrium E_1, the equilibrium price falls from $3.25 to $2.50, but the equilibrium quantity increases from 250,000 to 550,000 salmon. Notice that the equilibrium quantity demanded increased, even though the demand curve did not move.

In short, good weather conditions increased supply of the California commercial salmon. The result was a higher equilibrium quantity of salmon bought and sold in the market at a lower price.

Newspapers and the Internet

According to the Pew Research Center for People and the Press, increasingly more people, especially younger people, are obtaining their news from online and digital sources. The majority of U.S. adults now own smartphones or tablets, and most of those Americans say they use them in part to access the news. From 2004 to 2012, the share of Americans who reported obtaining their news from digital sources increased from 24% to 39%. How has this affected consumption of print news media, and radio and television news? **Figure 3.17** and the text below illustrates using the four-step analysis to answer this question.

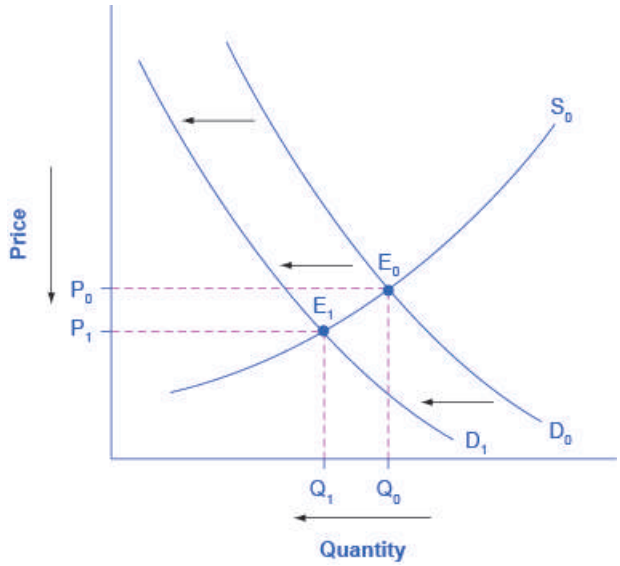


Figure 3.17 The Print News Market: A Four-Step Analysis A change in tastes from print news sources to digital sources results in a leftward shift in demand for the former. The result is a decrease in both equilibrium price and quantity.

Step 1. Develop a demand and supply model to think about what the market looked like before the event. The demand curve D_0 and the supply curve S_0 show the original relationships. In this case, we perform the analysis without

specific numbers on the price and quantity axis.

Step 2. Did the described change affect supply or demand? A change in tastes, from traditional news sources (print, radio, and television) to digital sources, caused a change in demand for the former.

Step 3. Was the effect on demand positive or negative? A shift to digital news sources will tend to mean a lower quantity demanded of traditional news sources at every given price, causing the demand curve for print and other traditional news sources to shift to the left, from D_0 to D_1.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_1) occurs at a lower quantity and a lower price than the original equilibrium (E_0).

The decline in print news reading predates 2004. Print newspaper circulation peaked in 1973 and has declined since then due to competition from television and radio news. In 1991, 55% of Americans indicated they received their news from print sources, while only 29% did so in 2012. Radio news has followed a similar path in recent decades, with the share of Americans obtaining their news from radio declining from 54% in 1991 to 33% in 2012. Television news has held its own over the last 15 years, with a market share staying in the mid to upper fifties. What does this suggest for the future, given that two-thirds of Americans under 30 years old say they do not obtain their news from television at all?

The Interconnections and Speed of Adjustment in Real Markets

In the real world, many factors that affect demand and supply can change all at once. For example, the demand for cars might increase because of rising incomes and population, and it might decrease because of rising gasoline prices (a complementary good). Likewise, the supply of cars might increase because of innovative new technologies that reduce the cost of car production, and it might decrease as a result of new government regulations requiring the installation of costly pollution-control technology.

Moreover, rising incomes and population or changes in gasoline prices will affect many markets, not just cars. How can an economist sort out all these interconnected events? The answer lies in the *ceteris paribus* assumption. Look at how each economic event affects each market, one event at a time, holding all else constant. Then combine the analyses to see the net effect.

A Combined Example

The U.S. Postal Service is facing difficult challenges. Compensation for postal workers tends to increase most years due to cost-of-living increases. At the same time, increasingly more people are using email, text, and other digital message forms such as Facebook and Twitter to communicate with friends and others. What does this suggest about the continued viability of the Postal Service? **Figure 3.18** and the text below illustrate this using the four-step analysis to answer this question.

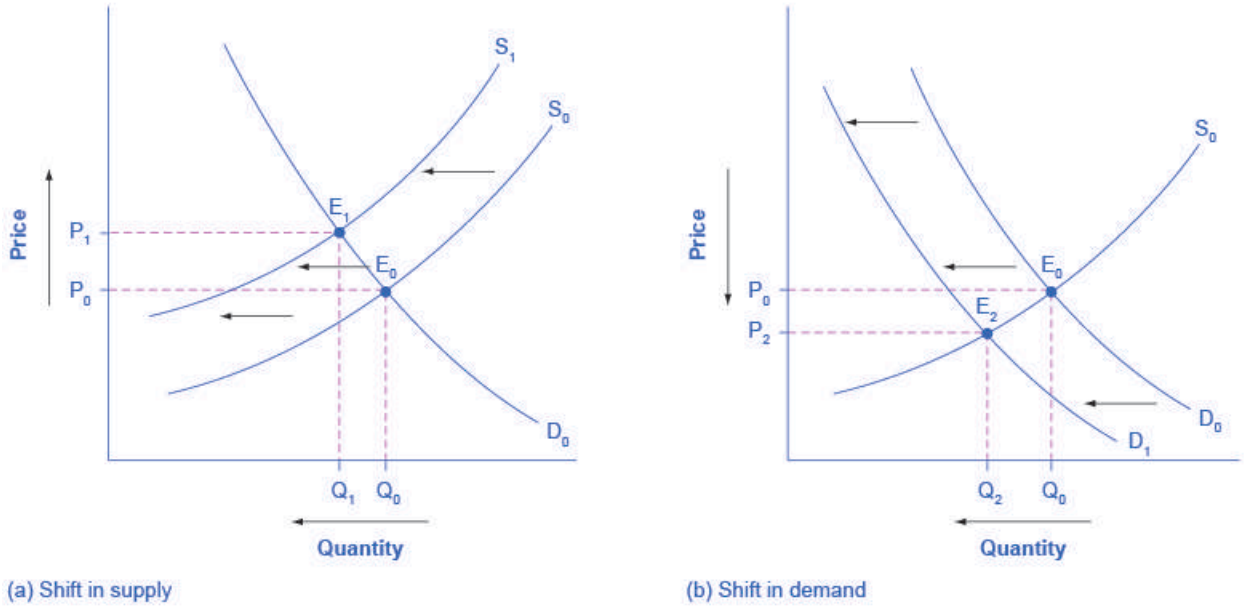


Figure 3.18 Higher Compensation for Postal Workers: A Four-Step Analysis (a) Higher labor compensation causes a leftward shift in the supply curve, a decrease in the equilibrium quantity, and an increase in the equilibrium price. (b) A change in tastes away from Postal Services causes a leftward shift in the demand curve, a decrease in the equilibrium quantity, and a decrease in the equilibrium price.

Since this problem involves two disturbances, we need two four-step analyses, the first to analyze the effects of higher compensation for postal workers, the second to analyze the effects of many people switching from "snail mail" to email and other digital messages.

Figure 3.18 (a) shows the shift in supply discussed in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for the U.S. Postal Service looked like before this scenario starts. The demand curve D_0 and the supply curve S_0 show the original relationships.

Step 2. Did the described change affect supply or demand? Labor compensation is a cost of production. A change in production costs caused a change in supply for the Postal Service.

Step 3. Was the effect on supply positive or negative? Higher labor compensation leads to a lower quantity supplied of postal services at every given price, causing the supply curve for postal services to shift to the left, from S_0 to S_1.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_1) occurs at a lower quantity and a higher price than the original equilibrium (E_0).

Figure 3.18 (b) shows the shift in demand in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for U.S. Postal Services looked like before this scenario starts. The demand curve D_0 and the supply curve S_0 show the original relationships. Note that this diagram is independent from the diagram in panel (a).

Step 2. Did the change described affect supply or demand? A change in tastes away from snail mail toward digital messages will cause a change in demand for the Postal Service.

Step 3. Was the effect on demand positive or negative? A change in tastes away from snailmail toward digital messages causes lower quantity demanded of postal services at every given price, causing the demand curve for postal services to shift to the left, from D_0 to D_1.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_2) occurs at a lower quantity and a lower price than the original equilibrium (E_0).

The final step in a scenario where both supply and demand shift is to combine the two individual analyses to determine what happens to the equilibrium quantity and price. Graphically, we superimpose the previous two diagrams one on top of the other, as in **Figure 3.19**.

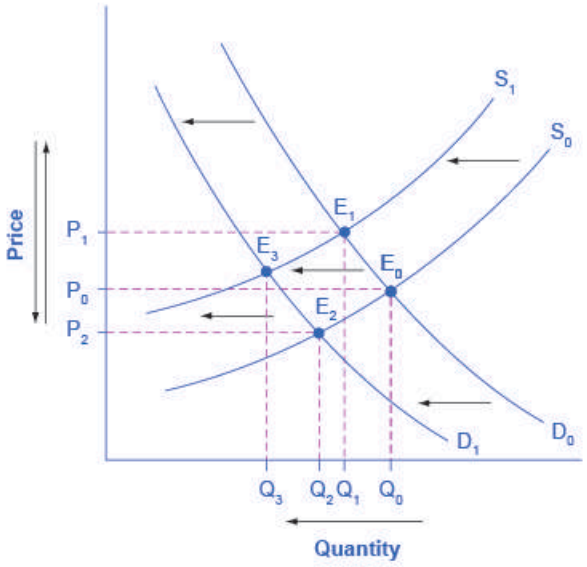


Figure 3.19 Combined Effect of Decreased Demand and Decreased Supply Supply and demand shifts cause changes in equilibrium price and quantity.

Following are the results:

Effect on Quantity: The effect of higher labor compensation on Postal Services because it raises the cost of production is to decrease the equilibrium quantity. The effect of a change in tastes away from snail mail is to decrease the equilibrium quantity. Since both shifts are to the left, the overall impact is a decrease in the equilibrium quantity of Postal Services (Q_3). This is easy to see graphically, since Q_3 is to the left of Q_0.

Effect on Price: The overall effect on price is more complicated. The effect of higher labor compensation on Postal Services, because it raises the cost of production, is to increase the equilibrium price. The effect of a change in tastes away from snail mail is to decrease the equilibrium price. Since the two effects are in opposite directions, unless we know the magnitudes of the two effects, the overall effect is unclear. This is not unusual. When both curves shift, typically we can determine the overall effect on price or on quantity, but not on both. In this case, we determined the overall effect on the equilibrium quantity, but not on the equilibrium price. In other cases, it might be the opposite.

The next Clear It Up feature focuses on the difference between shifts of supply or demand and movements along a curve.

Clear It Up

What is the difference between shifts of demand or supply versus movements along a demand or supply curve?

One common mistake in applying the demand and supply framework is to confuse the shift of a demand or a supply curve with movement along a demand or supply curve. As an example, consider a problem that asks whether a drought will increase or decrease the equilibrium quantity and equilibrium price of wheat. Lee, a student in an introductory economics class, might reason:

"Well, it is clear that a drought reduces supply, so I will shift back the supply curve, as in the shift from the original supply curve S_0 to S_1 on the diagram (Shift 1). The equilibrium moves from E_0 to E_1, the equilibrium quantity is lower and the equilibrium price is higher. Then, a higher price makes farmers more likely to supply the good, so the supply curve shifts right, as shows the shift from S_1 to S_2, shows on the diagram (Shift 2), so that the equilibrium now moves from E_1 to E_2. The higher price, however, also reduces demand and so causes demand to shift back, like the shift from the original demand curve, D_0 to D_1 on the diagram (labeled

Shift 3), and the equilibrium moves from E_2 to E_3."

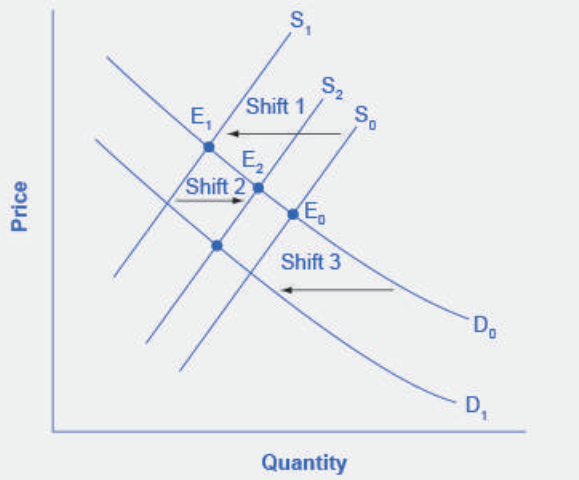


Figure 3.20 Shifts of Demand or Supply versus Movements along a Demand or Supply Curve A shift in one curve never causes a shift in the other curve. Rather, a shift in one curve causes a movement along the second curve.

At about this point, Lee suspects that this answer is headed down the wrong path. Think about what might be wrong with Lee's logic, and then read the answer that follows.

Answer: Lee's first step is correct: that is, a drought shifts back the supply curve of wheat and leads to a prediction of a lower equilibrium quantity and a higher equilibrium price. This corresponds to a movement along the original demand curve (D_0), from E_0 to E_1. The rest of Lee's argument is wrong, because it mixes up shifts in supply with quantity supplied, and shifts in demand with quantity demanded. A higher or lower price never shifts the supply curve, as suggested by the shift in supply from S_1 to S_2. Instead, a price change leads to a movement along a given supply curve. Similarly, a higher or lower price never shifts a demand curve, as suggested in the shift from D_0 to D_1. Instead, a price change leads to a movement along a given demand curve. Remember, a change in the price of a good never causes the demand or supply curve for that good to shift.

Think carefully about the timeline of events: What happens first, what happens next? What is cause, what is effect? If you keep the order right, you are more likely to get the analysis correct.

In the four-step analysis of how economic events affect equilibrium price and quantity, the movement from the old to the new equilibrium seems immediate. As a practical matter, however, prices and quantities often do not zoom straight to equilibrium. More realistically, when an economic event causes demand or supply to shift, prices and quantities set off in the general direction of equilibrium. Even as they are moving toward one new equilibrium, a subsequent change in demand or supply often pushes prices toward another equilibrium.

3.4 | Price Ceilings and Price Floors

By the end of this section, you will be able to:

- Explain price controls, price ceilings, and price floors
- Analyze demand and supply as a social adjustment mechanism

To this point in the chapter, we have been assuming that markets are free, that is, they operate with no government intervention. In this section, we will explore the outcomes, both anticipated and otherwise, when government does intervene in a market either to prevent the price of some good or service from rising "too high" or to prevent the price of some good or service from falling "too low".

Economists believe there are a small number of fundamental principles that explain how economic agents respond in different situations. Two of these principles, which we have already introduced, are the laws of demand and supply.

Governments can pass laws affecting market outcomes, but no law can negate these economic principles. Rather, the principles will become apparent in sometimes unexpected ways, which may undermine the intent of the government policy. This is one of the major conclusions of this section.

Controversy sometimes surrounds the prices and quantities established by demand and supply, especially for products that are considered necessities. In some cases, discontent over prices turns into public pressure on politicians, who may then pass legislation to prevent a certain price from climbing "too high" or falling "too low."

The demand and supply model shows how people and firms will react to the incentives that these laws provide to control prices, in ways that will often lead to undesirable consequences. Alternative policy tools can often achieve the desired goals of price control laws, while avoiding at least some of their costs and tradeoffs.

Price Ceilings

Laws that government enact to regulate prices are called **price controls**. Price controls come in two flavors. A **price ceiling** keeps a price from rising above a certain level (the "ceiling"), while a **price floor** keeps a price from falling below a given level (the "floor"). This section uses the demand and supply framework to analyze price ceilings. The next section discusses price floors.

A price ceiling is a legal maximum price that one pays for some good or service. A government imposes price ceilings in order to keep the price of some necessary good or service affordable. For example, in 2005 during Hurricane Katrina, the price of bottled water increased above $5 per gallon. As a result, many people called for price controls on bottled water to prevent the price from rising so high. In this particular case, the government did not impose a price ceiling, but there are other examples of where price ceilings did occur.

In many markets for goods and services, demanders outnumber suppliers. Consumers, who are also potential voters, sometimes unite behind a political proposal to hold down a certain price. In some cities, such as Albany, renters have pressed political leaders to pass rent control laws, a price ceiling that usually works by stating that landlords can raise rents by only a certain maximum percentage each year. Some of the best examples of rent control occur in urban areas such as New York, Washington D.C., or San Francisco.

Rent control becomes a politically hot topic when rents begin to rise rapidly. Everyone needs an affordable place to live. Perhaps a change in tastes makes a certain suburb or town a more popular place to live. Perhaps locally-based businesses expand, bringing higher incomes and more people into the area. Such changes can cause a change in the demand for rental housing, as **Figure 3.21** illustrates. The original equilibrium (E_0) lies at the intersection of supply curve S_0 and demand curve D_0, corresponding to an equilibrium price of $500 and an equilibrium quantity of 15,000 units of rental housing. The effect of greater income or a change in tastes is to shift the demand curve for rental housing to the right, as the data in **Table 3.7** shows and the shift from D_0 to D_1 on the graph. In this market, at the new equilibrium E_1, the price of a rental unit would rise to $600 and the equilibrium quantity would increase to 17,000 units.

Figure 3.21 A Price Ceiling Example—Rent Control The original intersection of demand and supply occurs at E_0. If demand shifts from D_0 to D_1, the new equilibrium would be at E_1—unless a price ceiling prevents the price from rising. If the price is not permitted to rise, the quantity supplied remains at 15,000. However, after the change in demand, the quantity demanded rises to 19,000, resulting in a shortage.

Price	Original Quantity Supplied	Original Quantity Demanded	New Quantity Demanded
$400	12,000	18,000	23,000
$500	15,000	15,000	19,000
$600	17,000	13,000	17,000
$700	19,000	11,000	15,000
$800	20,000	10,000	14,000

Table 3.7 Rent Control

Suppose that a city government passes a rent control law to keep the price at the original equilibrium of $500 for a typical apartment. In **Figure 3.21**, the horizontal line at the price of $500 shows the legally fixed maximum price set by the rent control law. However, the underlying forces that shifted the demand curve to the right are still there. At that price ($500), the quantity supplied remains at the same 15,000 rental units, but the quantity demanded is 19,000 rental units. In other words, the quantity demanded exceeds the quantity supplied, so there is a shortage of rental housing. One of the ironies of price ceilings is that while the price ceiling was intended to help renters, there are actually fewer apartments rented out under the price ceiling (15,000 rental units) than would be the case at the market rent of $600 (17,000 rental units).

Price ceilings do not simply benefit renters at the expense of landlords. Rather, some renters (or potential renters) lose their housing as landlords convert apartments to co-ops and condos. Even when the housing remains in the rental market, landlords tend to spend less on maintenance and on essentials like heating, cooling, hot water, and lighting. The first rule of economics is you do not get something for nothing—everything has an opportunity cost. Thus, if renters obtain “cheaper” housing than the market requires, they tend to also end up with lower quality housing.

Price ceilings are enacted in an attempt to keep prices low for those who need the product. However, when the market price is not allowed to rise to the equilibrium level, quantity demanded exceeds quantity supplied, and thus a shortage occurs. Those who manage to purchase the product at the lower price given by the price ceiling will benefit, but sellers of the product will suffer, along with those who are not able to purchase the product at all. Quality is also likely to deteriorate.

Price Floors

A price floor is the lowest price that one can legally pay for some good or service. Perhaps the best-known example

of a price floor is the minimum wage, which is based on the view that someone working full time should be able to afford a basic standard of living. The federal minimum wage in 2016 was $7.25 per hour, although some states and localities have a higher minimum wage. The federal minimum wage yields an annual income for a single person of $15,080, which is slightly higher than the Federal poverty line of $11,880. As the cost of living rises over time, the Congress periodically raises the federal minimum wage.

Price floors are sometimes called “price supports,” because they support a price by preventing it from falling below a certain level. Around the world, many countries have passed laws to create agricultural price supports. Farm prices and thus farm incomes fluctuate, sometimes widely. Even if, on average, farm incomes are adequate, some years they can be quite low. The purpose of price supports is to prevent these swings.

The most common way price supports work is that the government enters the market and buys up the product, adding to demand to keep prices higher than they otherwise would be. According to the Common Agricultural Policy reform passed in 2013, the European Union (EU) will spend about 60 billion euros per year, or 67 billion dollars per year (with the November 2016 exchange rate), or roughly 38% of the EU budget, on price supports for Europe’s farmers from 2014 to 2020.

Figure 3.22 illustrates the effects of a government program that assures a price above the equilibrium by focusing on the market for wheat in Europe. In the absence of government intervention, the price would adjust so that the quantity supplied would equal the quantity demanded at the equilibrium point E_0, with price P_0 and quantity Q_0. However, policies to keep prices high for farmers keeps the price above what would have been the market equilibrium level—the price Pf shown by the dashed horizontal line in the diagram. The result is a quantity supplied in excess of the quantity demanded (Qd). When quantity supplied exceeds quantity demanded, a surplus exists.

Economists estimate that the high-income areas of the world, including the United States, Europe, and Japan, spend roughly $1 billion per day in supporting their farmers. If the government is willing to purchase the excess supply (or to provide payments for others to purchase it), then farmers will benefit from the price floor, but taxpayers and consumers of food will pay the costs. Agricultural economists and policy makers have offered numerous proposals for reducing farm subsidies. In many countries, however, political support for subsidies for farmers remains strong. This is either because the population views this as supporting the traditional rural way of life or because of industry's lobbying power of the agro-business.

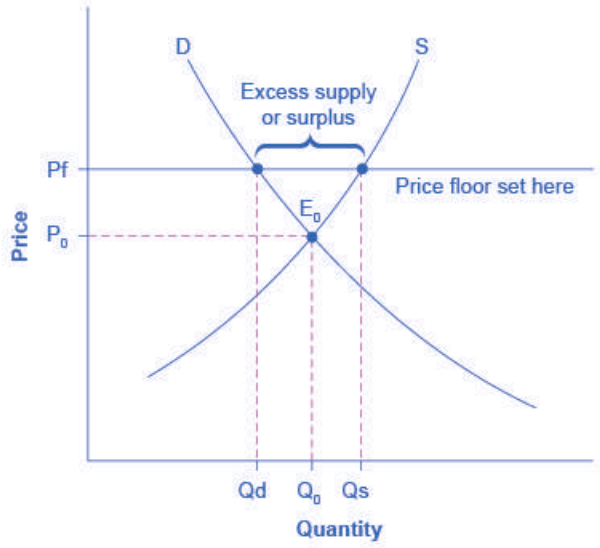


Figure 3.22 European Wheat Prices: A Price Floor Example The intersection of demand (D) and supply (S) would be at the equilibrium point E_0. However, a price floor set at Pf holds the price above E_0 and prevents it from falling. The result of the price floor is that the quantity supplied Qs exceeds the quantity demanded Qd. There is excess supply, also called a surplus.

3.5 | Demand, Supply, and Efficiency

By the end of this section, you will be able to:

- Contrast consumer surplus, producer surplus, and social surplus
- Explain why price floors and price ceilings can be inefficient
- Analyze demand and supply as a social adjustment mechanism

The familiar demand and supply diagram holds within it the concept of economic efficiency. One typical way that economists define efficiency is when it is impossible to improve the situation of one party without imposing a cost on another. Conversely, if a situation is inefficient, it becomes possible to benefit at least one party without imposing costs on others.

Efficiency in the demand and supply model has the same basic meaning: The economy is getting as much benefit as possible from its scarce resources and all the possible gains from trade have been achieved. In other words, the optimal amount of each good and service is produced and consumed.

Consumer Surplus, Producer Surplus, Social Surplus

Consider a market for tablet computers, as **Figure 3.23** shows. The equilibrium price is $80 and the equilibrium quantity is 28 million. To see the benefits to consumers, look at the segment of the demand curve above the equilibrium point and to the left. This portion of the demand curve shows that at least some demanders would have been willing to pay more than $80 for a tablet.

For example, point J shows that if the price were $90, 20 million tablets would be sold. Those consumers who would have been willing to pay $90 for a tablet based on the utility they expect to receive from it, but who were able to pay the equilibrium price of $80, clearly received a benefit beyond what they had to pay. Remember, the demand curve traces consumers' willingness to pay for different quantities. The amount that individuals would have been willing to pay, minus the amount that they actually paid, is called **consumer surplus**. Consumer surplus is the area labeled F—that is, the area above the market price and below the demand curve.

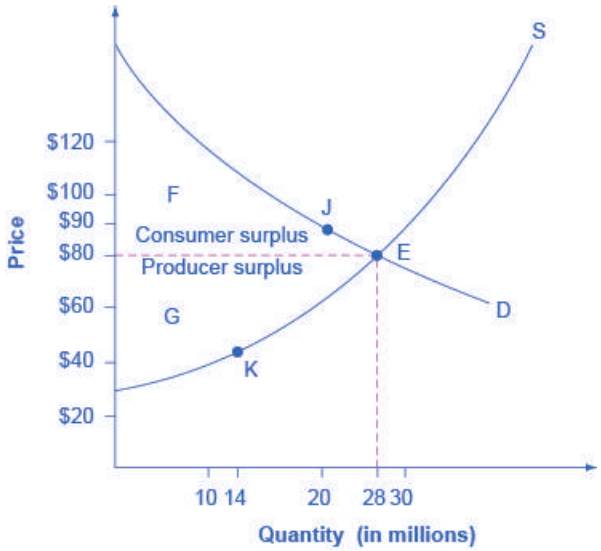


Figure 3.23 Consumer and Producer Surplus The somewhat triangular area labeled by F shows the area of consumer surplus, which shows that the equilibrium price in the market was less than what many of the consumers were willing to pay. Point J on the demand curve shows that, even at the price of $90, consumers would have been willing to purchase a quantity of 20 million. The somewhat triangular area labeled by G shows the area of producer surplus, which shows that the equilibrium price received in the market was more than what many of the producers were willing to accept for their products. For example, point K on the supply curve shows that at a price of $45, firms would have been willing to supply a quantity of 14 million.

The supply curve shows the quantity that firms are willing to supply at each price. For example, point K in **Figure**

3.23 illustrates that, at $45, firms would still have been willing to supply a quantity of 14 million. Those producers who would have been willing to supply the tablets at $45, but who were instead able to charge the equilibrium price of $80, clearly received an extra benefit beyond what they required to supply the product. The amount that a seller is paid for a good minus the seller's actual cost is called **producer surplus**. In Figure 3.23, producer surplus is the area labeled G—that is, the area between the market price and the segment of the supply curve below the equilibrium.

The sum of consumer surplus and producer surplus is **social surplus**, also referred to as **economic surplus** or **total surplus**. In Figure 3.23 we show social surplus as the area F + G. Social surplus is larger at equilibrium quantity and price than it would be at any other quantity. This demonstrates the economic efficiency of the market equilibrium. In addition, at the efficient level of output, it is impossible to produce greater consumer surplus without reducing producer surplus, and it is impossible to produce greater producer surplus without reducing consumer surplus.

Inefficiency of Price Floors and Price Ceilings

The imposition of a price floor or a price ceiling will prevent a market from adjusting to its equilibrium price and quantity, and thus will create an inefficient outcome. However, there is an additional twist here. Along with creating inefficiency, price floors and ceilings will also transfer some consumer surplus to producers, or some producer surplus to consumers.

Imagine that several firms develop a promising but expensive new drug for treating back pain. If this therapy is left to the market, the equilibrium price will be $600 per month and 20,000 people will use the drug, as shown in Figure 3.24 (a). The original level of consumer surplus is T + U and producer surplus is V + W + X. However, the government decides to impose a price ceiling of $400 to make the drug more affordable. At this price ceiling, firms in the market now produce only 15,000.

As a result, two changes occur. First, an inefficient outcome occurs and the total surplus of society is reduced. The loss in social surplus that occurs when the economy produces at an inefficient quantity is called **deadweight loss**. In a very real sense, it is like money thrown away that benefits no one. In Figure 3.24 (a), the deadweight loss is the area U + W. When deadweight loss exists, it is possible for both consumer and producer surplus to be higher, in this case because the price control is blocking some suppliers and demanders from transactions they would both be willing to make.

A second change from the price ceiling is that some of the producer surplus is transferred to consumers. After the price ceiling is imposed, the new consumer surplus is T + V, while the new producer surplus is X. In other words, the price ceiling transfers the area of surplus (V) from producers to consumers. Note that the gain to consumers is less than the loss to producers, which is just another way of seeing the deadweight loss.

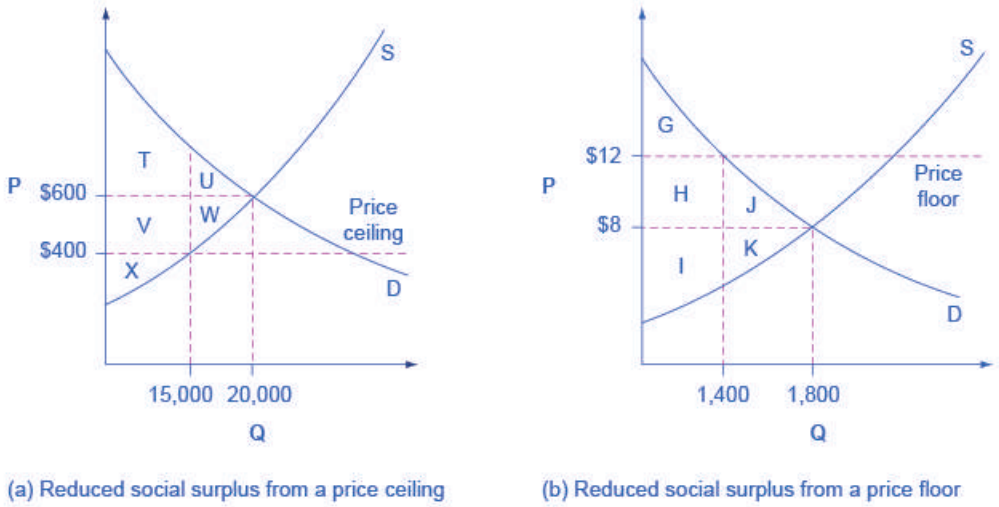


Figure 3.24 Efficiency and Price Floors and Ceilings (a) The original equilibrium price is $600 with a quantity of 20,000. Consumer surplus is T + U, and producer surplus is V + W + X. A price ceiling is imposed at $400, so firms in the market now produce only a quantity of 15,000. As a result, the new consumer surplus is T + V, while the new producer surplus is X. (b) The original equilibrium is $8 at a quantity of 1,800. Consumer surplus is G + H + J, and producer surplus is I + K. A price floor is imposed at $12, which means that quantity demanded falls to 1,400. As a result, the new consumer surplus is G, and the new producer surplus is H + I.

Figure 3.24 (b) shows a price floor example using a string of struggling movie theaters, all in the same city. The current equilibrium is $8 per movie ticket, with 1,800 people attending movies. The original consumer surplus is G + H + J, and producer surplus is I + K. The city government is worried that movie theaters will go out of business, reducing the entertainment options available to citizens, so it decides to impose a price floor of $12 per ticket. As a result, the quantity demanded of movie tickets falls to 1,400. The new consumer surplus is G, and the new producer surplus is H + I. In effect, the price floor causes the area H to be transferred from consumer to producer surplus, but also causes a deadweight loss of J + K.

This analysis shows that a price ceiling, like a law establishing rent controls, will transfer some producer surplus to consumers—which helps to explain why consumers often favor them. Conversely, a price floor like a guarantee that farmers will receive a certain price for their crops will transfer some consumer surplus to producers, which explains why producers often favor them. However, both price floors and price ceilings block some transactions that buyers and sellers would have been willing to make, and creates deadweight loss. Removing such barriers, so that prices and quantities can adjust to their equilibrium level, will increase the economy's social surplus.

Demand and Supply as a Social Adjustment Mechanism

The demand and supply model emphasizes that prices are not set only by demand or only by supply, but by the interaction between the two. In 1890, the famous economist Alfred Marshall wrote that asking whether supply or demand determined a price was like arguing "whether it is the upper or the under blade of a pair of scissors that cuts a piece of paper." The answer is that both blades of the demand and supply scissors are always involved.

The adjustments of equilibrium price and quantity in a market-oriented economy often occur without much government direction or oversight. If the coffee crop in Brazil suffers a terrible frost, then the supply curve of coffee shifts to the left and the price of coffee rises. Some people—call them the coffee addicts—continue to drink coffee and pay the higher price. Others switch to tea or soft drinks. No government commission is needed to figure out how to adjust coffee prices, which companies will be allowed to process the remaining supply, which supermarkets in which cities will get how much coffee to sell, or which consumers will ultimately be allowed to drink the brew. Such adjustments in response to price changes happen all the time in a market economy, often so smoothly and rapidly that we barely notice them.

Think for a moment of all the seasonal foods that are available and inexpensive at certain times of the year, like fresh corn in midsummer, but more expensive at other times of the year. People alter their diets and restaurants alter their menus in response to these fluctuations in prices without fuss or fanfare. For both the U.S. economy and the world

economy as a whole, markets—that is, demand and supply—are the primary social mechanism for answering the basic questions about what is produced, how it is produced, and for whom it is produced.

Bring it Home

Why Can We Not Get Enough of Organic?

Organic food is grown without synthetic pesticides, chemical fertilizers or genetically modified seeds. In recent decades, the demand for organic products has increased dramatically. The Organic Trade Association reported sales increased from $1 billion in 1990 to $35.1 billion in 2013, more than 90% of which were sales of food products.

Why, then, are organic foods more expensive than their conventional counterparts? The answer is a clear application of the theories of supply and demand. As people have learned more about the harmful effects of chemical fertilizers, growth hormones, pesticides and the like from large-scale factory farming, our tastes and preferences for safer, organic foods have increased. This change in tastes has been reinforced by increases in income, which allow people to purchase pricier products, and has made organic foods more mainstream. This has led to an increased demand for organic foods. Graphically, the demand curve has shifted right, and we have moved up the supply curve as producers have responded to the higher prices by supplying a greater quantity.

In addition to the movement along the supply curve, we have also had an increase in the number of farmers converting to organic farming over time. This is represented by a shift to the right of the supply curve. Since both demand and supply have shifted to the right, the resulting equilibrium quantity of organic foods is definitely higher, but the price will only fall when the increase in supply is larger than the increase in demand. We may need more time before we see lower prices in organic foods. Since the production costs of these foods may remain higher than conventional farming, because organic fertilizers and pest management techniques are more expensive, they may never fully catch up with the lower prices of non-organic foods.

As a final, specific example: The Environmental Working Group's "Dirty Dozen" list of fruits and vegetables, which test high for pesticide residue even after washing, was released in April 2013. The inclusion of strawberries on the list has led to an increase in demand for organic strawberries, resulting in both a higher equilibrium price and quantity of sales.

KEY TERMS

ceteris paribus other things being equal

complements goods that are often used together so that consumption of one good tends to enhance consumption of the other

consumer surplus the extra benefit consumers receive from buying a good or service, measured by what the individuals would have been willing to pay minus the amount that they actually paid

deadweight loss the loss in social surplus that occurs when a market produces an inefficient quantity

demand the relationship between price and the quantity demanded of a certain good or service

demand curve a graphic representation of the relationship between price and quantity demanded of a certain good or service, with quantity on the horizontal axis and the price on the vertical axis

demand schedule a table that shows a range of prices for a certain good or service and the quantity demanded at each price

economic surplus see social surplus

equilibrium the situation where quantity demanded is equal to the quantity supplied; the combination of price and quantity where there is no economic pressure from surpluses or shortages that would cause price or quantity to change

equilibrium price the price where quantity demanded is equal to quantity supplied

equilibrium quantity the quantity at which quantity demanded and quantity supplied are equal for a certain price level

excess demand at the existing price, the quantity demanded exceeds the quantity supplied; also called a shortage

excess supply at the existing price, quantity supplied exceeds the quantity demanded; also called a surplus

factors of production the resources such as labor, materials, and machinery that are used to produce goods and services; also called inputs

inferior good a good in which the quantity demanded falls as income rises, and in which quantity demanded rises and income falls

inputs the resources such as labor, materials, and machinery that are used to produce goods and services; also called factors of production

law of demand the common relationship that a higher price leads to a lower quantity demanded of a certain good or service and a lower price leads to a higher quantity demanded, while all other variables are held constant

law of supply the common relationship that a higher price leads to a greater quantity supplied and a lower price leads to a lower quantity supplied, while all other variables are held constant

normal good a good in which the quantity demanded rises as income rises, and in which quantity demanded falls as income falls

price what a buyer pays for a unit of the specific good or service

price ceiling a legal maximum price

price control government laws to regulate prices instead of letting market forces determine prices

price floor a legal minimum price

producer surplus the extra benefit producers receive from selling a good or service, measured by the price the producer actually received minus the price the producer would have been willing to accept

quantity demanded the total number of units of a good or service consumers are willing to purchase at a given price

quantity supplied the total number of units of a good or service producers are willing to sell at a given price

shift in demand when a change in some economic factor (other than price) causes a different quantity to be demanded at every price

shift in supply when a change in some economic factor (other than price) causes a different quantity to be supplied at every price

shortage at the existing price, the quantity demanded exceeds the quantity supplied; also called excess demand

social surplus the sum of consumer surplus and producer surplus

substitute a good that can replace another to some extent, so that greater consumption of one good can mean less of the other

supply the relationship between price and the quantity supplied of a certain good or service

supply curve a line that shows the relationship between price and quantity supplied on a graph, with quantity supplied on the horizontal axis and price on the vertical axis

supply schedule a table that shows a range of prices for a good or service and the quantity supplied at each price

surplus at the existing price, quantity supplied exceeds the quantity demanded; also called excess supply

total surplus see social surplus

KEY CONCEPTS AND SUMMARY

3.1 Demand, Supply, and Equilibrium in Markets for Goods and Services

A demand schedule is a table that shows the quantity demanded at different prices in the market. A demand curve shows the relationship between quantity demanded and price in a given market on a graph. The law of demand states that a higher price typically leads to a lower quantity demanded.

A supply schedule is a table that shows the quantity supplied at different prices in the market. A supply curve shows the relationship between quantity supplied and price on a graph. The law of supply says that a higher price typically leads to a higher quantity supplied.

The equilibrium price and equilibrium quantity occur where the supply and demand curves cross. The equilibrium occurs where the quantity demanded is equal to the quantity supplied. If the price is below the equilibrium level, then the quantity demanded will exceed the quantity supplied. Excess demand or a shortage will exist. If the price is above the equilibrium level, then the quantity supplied will exceed the quantity demanded. Excess supply or a surplus will exist. In either case, economic pressures will push the price toward the equilibrium level.

3.2 Shifts in Demand and Supply for Goods and Services

Economists often use the *ceteris paribus* or “other things being equal” assumption: while examining the economic impact of one event, all other factors remain unchanged for analysis purposes. Factors that can shift the demand curve for goods and services, causing a different quantity to be demanded at any given price, include changes in tastes, population, income, prices of substitute or complement goods, and expectations about future conditions and prices. Factors that can shift the supply curve for goods and services, causing a different quantity to be supplied at any given price, include input prices, natural conditions, changes in technology, and government taxes, regulations, or subsidies.

3.3 Changes in Equilibrium Price and Quantity: The Four-Step Process

When using the supply and demand framework to think about how an event will affect the equilibrium price and quantity, proceed through four steps: (1) sketch a supply and demand diagram to think about what the market looked like before the event; (2) decide whether the event will affect supply or demand; (3) decide whether the effect on supply or demand is negative or positive, and draw the appropriate shifted supply or demand curve; (4) compare the new equilibrium price and quantity to the original ones.

3.4 Price Ceilings and Price Floors

Price ceilings prevent a price from rising above a certain level. When a price ceiling is set below the equilibrium price, quantity demanded will exceed quantity supplied, and excess demand or shortages will result. Price floors prevent a price from falling below a certain level. When a price floor is set above the equilibrium price, quantity supplied will exceed quantity demanded, and excess supply or surpluses will result. Price floors and price ceilings often lead to unintended consequences.

3.5 Demand, Supply, and Efficiency

Consumer surplus is the gap between the price that consumers are willing to pay, based on their preferences, and the market equilibrium price. Producer surplus is the gap between the price for which producers are willing to sell a product, based on their costs, and the market equilibrium price. Social surplus is the sum of consumer surplus and producer surplus. Total surplus is larger at the equilibrium quantity and price than it will be at any other quantity and price. Deadweight loss is loss in total surplus that occurs when the economy produces at an inefficient quantity.

SELF-CHECK QUESTIONS

1. Review **Figure 3.4**. Suppose the price of gasoline is $1.60 per gallon. Is the quantity demanded higher or lower than at the equilibrium price of $1.40 per gallon? What about the quantity supplied? Is there a shortage or a surplus in the market? If so, how much?

2. Why do economists use the *ceteris paribus* assumption?

3. In an analysis of the market for paint, an economist discovers the facts listed below. State whether each of these changes will affect supply or demand, and in what direction.

a. There have recently been some important cost-saving inventions in the technology for making paint.
b. Paint is lasting longer, so that property owners need not repaint as often.
c. Because of severe hailstorms, many people need to repaint now.
d. The hailstorms damaged several factories that make paint, forcing them to close down for several months.

4. Many changes are affecting the market for oil. Predict how each of the following events will affect the equilibrium price and quantity in the market for oil. In each case, state how the event will affect the supply and demand diagram. Create a sketch of the diagram if necessary.

a. Cars are becoming more fuel efficient, and therefore get more miles to the gallon.
b. The winter is exceptionally cold.
c. A major discovery of new oil is made off the coast of Norway.
d. The economies of some major oil-using nations, like Japan, slow down.
e. A war in the Middle East disrupts oil-pumping schedules.
f. Landlords install additional insulation in buildings.
g. The price of solar energy falls dramatically.
h. Chemical companies invent a new, popular kind of plastic made from oil.

5. Let’s think about the market for air travel. From August 2014 to January 2015, the price of jet fuel decreased roughly 47%. Using the four-step analysis, how do you think this fuel price decrease affected the equilibrium price and quantity of air travel?

6. A tariff is a tax on imported goods. Suppose the U.S. government cuts the tariff on imported flat screen televisions. Using the four-step analysis, how do you think the tariff reduction will affect the equilibrium price and quantity of flat screen TVs?

7. What is the effect of a price ceiling on the quantity demanded of the product? What is the effect of a price ceiling on the quantity supplied? Why exactly does a price ceiling cause a shortage?

8. Does a price ceiling change the equilibrium price?

9. What would be the impact of imposing a price floor below the equilibrium price?

10. Does a price ceiling increase or decrease the number of transactions in a market? Why? What about a price floor?

11. If a price floor benefits producers, why does a price floor reduce social surplus?

REVIEW QUESTIONS

12. What determines the level of prices in a market?

13. What does a downward-sloping demand curve mean about how buyers in a market will react to a higher price?

14. Will demand curves have the same exact shape in all markets? If not, how will they differ?

15. Will supply curves have the same shape in all markets? If not, how will they differ?

16. What is the relationship between quantity demanded and quantity supplied at equilibrium? What is the relationship when there is a shortage? What is the relationship when there is a surplus?

17. How can you locate the equilibrium point on a demand and supply graph?

18. If the price is above the equilibrium level, would you predict a surplus or a shortage? If the price is below the equilibrium level, would you predict a surplus or a shortage? Why?

19. When the price is above the equilibrium, explain how market forces move the market price to equilibrium. Do the same when the price is below the equilibrium.

20. What is the difference between the demand and the quantity demanded of a product, say milk? Explain in words and show the difference on a graph with a demand curve for milk.

21. What is the difference between the supply and the quantity supplied of a product, say milk? Explain in words and show the difference on a graph with the supply curve for milk.

22. When analyzing a market, how do economists deal with the problem that many factors that affect the market are changing at the same time?

23. Name some factors that can cause a shift in the demand curve in markets for goods and services.

24. Name some factors that can cause a shift in the supply curve in markets for goods and services.

25. How does one analyze a market where both demand and supply shift?

26. What causes a movement along the demand curve? What causes a movement along the supply curve?

27. Does a price ceiling attempt to make a price higher or lower?

28. How does a price ceiling set below the equilibrium level affect quantity demanded and quantity supplied?

29. Does a price floor attempt to make a price higher or lower?

30. How does a price floor set above the equilibrium level affect quantity demanded and quantity supplied?

31. What is consumer surplus? How is it illustrated on a demand and supply diagram?

32. What is producer surplus? How is it illustrated on a demand and supply diagram?

33. What is total surplus? How is it illustrated on a demand and supply diagram?

34. What is the relationship between total surplus and economic efficiency?

35. What is deadweight loss?

CRITICAL THINKING QUESTIONS

36. Review **Figure 3.4**. Suppose the government decided that, since gasoline is a necessity, its price should be legally capped at $1.30 per gallon. What do you anticipate would be the outcome in the gasoline market?

37. Explain why the following statement is false: "In the goods market, no buyer would be willing to pay more than the equilibrium price."

38. Explain why the following statement is false: "In the goods market, no seller would be willing to sell for less than the equilibrium price."

39. Consider the demand for hamburgers. If the price of a substitute good (for example, hot dogs) increases and the price of a complement good (for example, hamburger buns) increases, can you tell for sure what will happen to the demand for hamburgers? Why or why not? Illustrate your answer with a graph.

40. How do you suppose the demographics of an aging population of "Baby Boomers" in the United States will affect the demand for milk? Justify your answer.

41. We know that a change in the price of a product causes a movement along the demand curve. Suppose consumers believe that prices will be rising in the future. How will that affect demand for the product in the present? Can you show this graphically?

42. Suppose there is a soda tax to curb obesity. What should a reduction in the soda tax do to the supply of sodas and to the equilibrium price and quantity? Can you show this graphically? *Hint*: Assume that the soda tax is collected from the sellers.

43. Use the four-step process to analyze the impact of the advent of the iPod (or other portable digital music players) on the equilibrium price and quantity of the Sony Walkman (or other portable audio cassette players).

44. Use the four-step process to analyze the impact of a reduction in tariffs on imports of iPods on the equilibrium price and quantity of Sony Walkman-type products.

45. Suppose both of these events took place at the same time. Combine your analyses of the impacts of the iPod and the tariff reduction to determine the likely impact on the equilibrium price and quantity of Sony Walkman-type products. Show your answer graphically.

46. Most government policy decisions have winners and losers. What are the effects of raising the minimum wage? It is more complex than simply producers lose and workers gain. Who are the winners and who are the losers, and what exactly do they win and lose? To what extent does the policy change achieve its goals?

47. Agricultural price supports result in governments holding large inventories of agricultural products. Why do you think the government cannot simply give the products away to poor people?

48. Can you propose a policy that would induce the market to supply more rental housing units?

49. What term would an economist use to describe what happens when a shopper gets a "good deal" on a product?

50. Explain why voluntary transactions improve social welfare.

51. Why would a free market never operate at a quantity greater than the equilibrium quantity? *Hint:* What would be required for a transaction to occur at that quantity?

PROBLEMS

52. Review **Figure 3.4** again. Suppose the price of gasoline is $1.00. Will the quantity demanded be lower or higher than at the equilibrium price of $1.40 per gallon? Will the quantity supplied be lower or higher? Is there a shortage or a surplus in the market? If so, of how much?

53. Table 3.8 shows information on the demand and supply for bicycles, where the quantities of bicycles are measured in thousands.

Price	Qd	Qs
$120	50	36
$150	40	40
$180	32	48
$210	28	56
$240	24	70

Table 3.8

a. What is the quantity demanded and the quantity supplied at a price of $210?
b. At what price is the quantity supplied equal to 48,000?
c. Graph the demand and supply curve for bicycles. How can you determine the equilibrium price and quantity from the graph? How can you determine the equilibrium price and quantity from the table? What are the equilibrium price and equilibrium quantity?
d. If the price was $120, what would the quantities demanded and supplied be? Would a shortage or surplus exist? If so, how large would the shortage or surplus be?

54. The computer market in recent years has seen many more computers sell at much lower prices. What shift in demand or supply is most likely to explain this outcome? Sketch a demand and supply diagram and explain your reasoning for each.

a. A rise in demand
b. A fall in demand
c. A rise in supply
d. A fall in supply

55. Table 3.9 illustrates the market's demand and supply for cheddar cheese. Graph the data and find the equilibrium. Next, create a table showing the change in quantity demanded or quantity supplied, and a graph of the new equilibrium, in each of the following situations:

a. The price of milk, a key input for cheese production, rises, so that the supply decreases by 80 pounds at every price.
b. A new study says that eating cheese is good for your health, so that demand increases by 20% at every price.

Price per Pound	Qd	Qs
$3.00	750	540
$3.20	700	600
$3.40	**650**	**650**
$3.60	620	700
$3.80	600	720
$4.00	590	730

Table 3.9

56. Table 3.10 shows the supply and demand for movie tickets in a city. Graph demand and supply and identify the equilibrium. Then calculate in a table and graph the effect of the following two changes.

a. Three new nightclubs open. They offer decent bands and have no cover charge, but make their money by selling food and drink. As a result, demand for movie tickets falls by six units at every price.
b. The city eliminates a tax that it placed on all local entertainment businesses. The result is that the quantity supplied of movies at any given price increases by 10%.

Price per Pound	Qd	Qs
$5.00	26	16
$6.00	24	18
$7.00	22	20
$8.00	21	21
$9.00	20	22

Table 3.10

57. A low-income country decides to set a price ceiling on bread so it can make sure that bread is affordable to the poor.Table 3.11 provides the conditions of demand and supply. What are the equilibrium price and equilibrium quantity before the price ceiling? What will the excess demand or the shortage (that is, quantity demanded minus quantity supplied) be if the price ceiling is set at $2.40? At $2.00? At $3.60?

Price	Qd	Qs
$1.60	9,000	5,000
$2.00	8,500	5,500
$2.40	8,000	6,400
$2.80	**7,500**	**7,500**
$3.20	7,000	9,000
$3.60	6,500	11,000
$4.00	6,000	15,000

Table 3.11

4 | Labor and Financial Markets

Figure 4.1 People often think of demand and supply in relation to goods, but labor markets, such as the nursing profession, can also apply to this analysis. (Credit: modification of work by "Fotos GOVBA"/Flickr Creative Commons)

Bring it Home

Baby Boomers Come of Age

The Census Bureau reports that as of 2013, 20% of the U.S. population was over 60 years old, which means that almost 63 million people are reaching an age when they will need increased medical care.

The baby boomer population, the group born between 1946 and 1964, is comprised of approximately 74 million people who have just reached retirement age. As this population grows older, they will be faced with common healthcare issues such as heart conditions, arthritis, and Alzheimer's that may require hospitalization, long-term, or at-home nursing care. Aging baby boomers and advances in life-saving and life-extending technologies will increase the demand for healthcare and nursing. Additionally, the Affordable Care Act, which expands access to healthcare for millions of Americans, has further increase the demand, although with the election of Donald J. Trump, this increase may not be sustained.

According to the Bureau of Labor Statistics, registered nursing jobs are expected to increase by 16% between 2014 and 2024. The median annual wage of $67,490 (in 2015) is also expected to increase. The BLS forecasts that 439,000 new nurses will be in demand by 2022.

These data tell us, as economists, that the market for healthcare professionals, and nurses in particular, will face several challenges. Our study of supply and demand will help us to analyze what might happen in the

labor market for nursing and other healthcare professionals, as we will discuss in the second half of this case at the end of the chapter.

Introduction to Labor and Financial Markets

In this chapter, you will learn about:

- Demand and Supply at Work in Labor Markets
- Demand and Supply in Financial Markets
- The Market System as an Efficient Mechanism for Information

The theories of supply and demand do not apply just to markets for goods. They apply to any market, even markets for things we may not think of as goods and services like labor and financial services. Labor markets are markets for employees or jobs. Financial services markets are markets for saving or borrowing.

When we think about demand and supply curves in goods and services markets, it is easy to picture the demanders and suppliers: businesses produce the products and households buy them. Who are the demanders and suppliers in labor and financial service markets? In labor markets job seekers (individuals) are the suppliers of labor, while firms and other employers who hire labor are the demanders for labor. In financial markets, any individual or firm who saves contributes to the supply of money, and any who borrows (person, firm, or government) contributes to the demand for money.

As a college student, you most likely participate in both labor and financial markets. Employment is a fact of life for most college students: According to the National Center for Educational Statistics, in 2013 40% of full-time college students and 76% of part-time college students were employed. Most college students are also heavily involved in financial markets, primarily as borrowers. Among full-time students, about half take out a loan to help finance their education each year, and those loans average about $6,000 per year. Many students also borrow for other expenses, like purchasing a car. As this chapter will illustrate, we can analyze labor markets and financial markets with the same tools we use to analyze demand and supply in the goods markets.

4.1 | Demand and Supply at Work in Labor Markets

By the end of this section, you will be able to:

- Predict shifts in the demand and supply curves of the labor market
- Explain the impact of new technology on the demand and supply curves of the labor market
- Explain price floors in the labor market such as minimum wage or a living wage

Markets for labor have demand and supply curves, just like markets for goods. The law of demand applies in labor markets this way: A higher **salary** or **wage**—that is, a higher price in the labor market—leads to a decrease in the quantity of labor demanded by employers, while a lower salary or wage leads to an increase in the quantity of labor demanded. The law of supply functions in labor markets, too: A higher price for labor leads to a higher quantity of labor supplied; a lower price leads to a lower quantity supplied.

Equilibrium in the Labor Market

In 2015, about 35,000 registered nurses worked in the Minneapolis-St. Paul-Bloomington, Minnesota-Wisconsin metropolitan area, according to the BLS. They worked for a variety of employers: hospitals, doctors' offices, schools, health clinics, and nursing homes. **Figure 4.2** illustrates how demand and supply determine equilibrium in this labor market. The demand and supply schedules in **Table 4.1** list the quantity supplied and quantity demanded of nurses at different salaries.

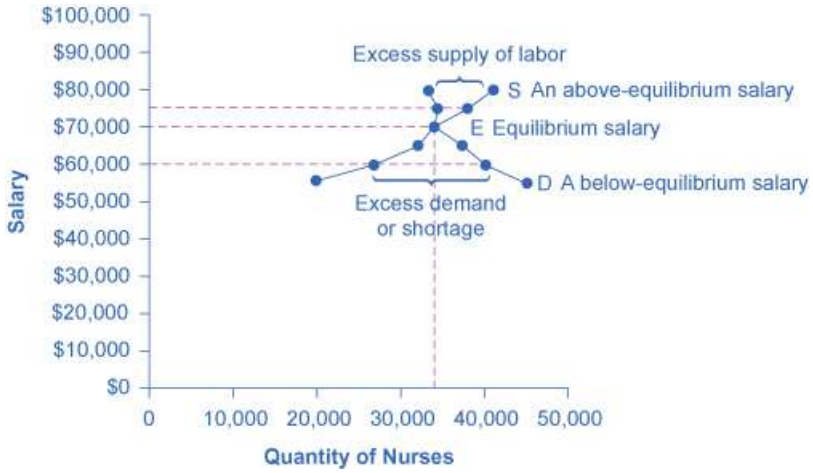


Figure 4.2 Labor Market Example: Demand and Supply for Nurses in Minneapolis-St. Paul-Bloomington The demand curve (D) of those employers who want to hire nurses intersects with the supply curve (S) of those who are qualified and willing to work as nurses at the equilibrium point (E). The equilibrium salary is $70,000 and the equilibrium quantity is 34,000 nurses. At an above-equilibrium salary of $75,000, quantity supplied increases to 38,000, but the quantity of nurses demanded at the higher pay declines to 33,000. At this above-equilibrium salary, an excess supply or surplus of nurses would exist. At a below-equilibrium salary of $60,000, quantity supplied declines to 27,000, while the quantity demanded at the lower wage increases to 40,000 nurses. At this below-equilibrium salary, excess demand or a shortage exists.

Annual Salary	Quantity Demanded	Quantity Supplied
$55,000	45,000	20,000
$60,000	40,000	27,000
$65,000	37,000	31,000
$70,000	34,000	34,000
$75,000	33,000	38,000
$80,000	32,000	41,000

Table 4.1 Demand and Supply of Nurses in Minneapolis-St. Paul-Bloomington

The horizontal axis shows the quantity of nurses hired. In this example we measure labor by number of workers, but another common way to measure the quantity of labor is by the number of hours worked. The vertical axis shows the price for nurses' labor—that is, how much they are paid. In the real world, this "price" would be total labor compensation: salary plus benefits. It is not obvious, but benefits are a significant part (as high as 30 percent) of labor compensation. In this example we measure the price of labor by salary on an annual basis, although in other cases we could measure the price of labor by monthly or weekly pay, or even the wage paid per hour. As the salary for nurses rises, the quantity demanded will fall. Some hospitals and nursing homes may reduce the number of nurses they hire, or they may lay off some of their existing nurses, rather than pay them higher salaries. Employers who face higher nurses' salaries may also try to replace some nursing functions by investing in physical equipment, like computer monitoring and diagnostic systems to monitor patients, or by using lower-paid health care aides to reduce the number of nurses they need.

As the salary for nurses rises, the quantity supplied will rise. If nurses' salaries in Minneapolis-St. Paul-Bloomington are higher than in other cities, more nurses will move to Minneapolis-St. Paul-Bloomington to find jobs, more people will be willing to train as nurses, and those currently trained as nurses will be more likely to pursue nursing as a full-time job. In other words, there will be more nurses looking for jobs in the area.

At **equilibrium**, the quantity supplied and the quantity demanded are equal. Thus, every employer who wants to hire a nurse at this equilibrium wage can find a willing worker, and every nurse who wants to work at this equilibrium salary can find a job. In **Figure 4.2**, the supply curve (S) and demand curve (D) intersect at the equilibrium point (E). The equilibrium quantity of nurses in the Minneapolis-St. Paul-Bloomington area is 34,000, and the equilibrium salary is $70,000 per year. This example simplifies the nursing market by focusing on the "average" nurse. In reality, of course, the market for nurses actually comprises many smaller markets, like markets for nurses with varying degrees of experience and credentials. Many markets contain closely related products that differ in quality. For instance, even a simple product like gasoline comes in regular, premium, and super-premium, each with a different price. Even in such cases, discussing the average price of gasoline, like the average salary for nurses, can still be useful because it reflects what is happening in most of the submarkets.

When the price of labor is not at the equilibrium, economic incentives tend to move salaries toward the equilibrium. For example, if salaries for nurses in Minneapolis-St. Paul-Bloomington were above the equilibrium at $75,000 per year, then 38,000 people want to work as nurses, but employers want to hire only 33,000 nurses. At that above-equilibrium salary, excess supply or a surplus results. In a situation of excess supply in the **labor market**, with many applicants for every job opening, employers will have an incentive to offer lower wages than they otherwise would have. Nurses' salary will move down toward equilibrium.

In contrast, if the salary is below the equilibrium at, say, $60,000 per year, then a situation of excess demand or a shortage arises. In this case, employers encouraged by the relatively lower wage want to hire 40,000 nurses, but only 27,000 individuals want to work as nurses at that salary in Minneapolis-St. Paul-Bloomington. In response to the shortage, some employers will offer higher pay to attract the nurses. Other employers will have to match the higher pay to keep their own employees. The higher salaries will encourage more nurses to train or work in Minneapolis-St. Paul-Bloomington. Again, price and quantity in the labor market will move toward equilibrium.

Shifts in Labor Demand

The demand curve for labor shows the quantity of labor employers wish to hire at any given salary or wage rate, under the *ceteris paribus* assumption. A change in the wage or salary will result in a change in the quantity demanded of labor. If the wage rate increases, employers will want to hire fewer employees. The quantity of labor demanded will decrease, and there will be a movement upward along the demand curve. If the wages and salaries decrease, employers are more likely to hire a greater number of workers. The quantity of labor demanded will increase, resulting in a downward movement along the demand curve.

Shifts in the demand curve for labor occur for many reasons. One key reason is that the demand for labor is based on the demand for the good or service that is produced. For example, the more new automobiles consumers demand, the greater the number of workers automakers will need to hire. Therefore the demand for labor is called a "derived demand." Here are some examples of derived demand for labor:

- The demand for chefs is dependent on the demand for restaurant meals.
- The demand for pharmacists is dependent on the demand for prescription drugs.
- The demand for attorneys is dependent on the demand for legal services.

As the demand for the goods and services increases, the demand for labor will increase, or shift to the right, to meet employers' production requirements. As the demand for the goods and services decreases, the demand for labor will decrease, or shift to the left. **Table 4.2** shows that in addition to the derived demand for labor, demand can also increase or decrease (shift) in response to several factors.

Factors	Results
Demand for Output	When the demand for the good produced (output) increases, both the output price and profitability increase. As a result, producers demand more labor to ramp up production.

Table 4.2 Factors That Can Shift Demand

Factors	Results
Education and Training	A well-trained and educated workforce causes an increase in the demand for that labor by employers. Increased levels of productivity within the workforce will cause the demand for labor to shift to the right. If the workforce is not well-trained or educated, employers will not hire from within that labor pool, since they will need to spend a significant amount of time and money training that workforce. Demand for such will shift to the left.
Technology	Technology changes can act as either substitutes for or complements to labor. When technology acts as a substitute, it replaces the need for the number of workers an employer needs to hire. For example, word processing decreased the number of typists needed in the workplace. This shifted the demand curve for typists left. An increase in the availability of certain technologies may increase the demand for labor. Technology that acts as a complement to labor will increase the demand for certain types of labor, resulting in a rightward shift of the demand curve. For example, the increased use of word processing and other software has increased the demand for information technology professionals who can resolve software and hardware issues related to a firm's network. More and better technology will increase demand for skilled workers who know how to use technology to enhance workplace productivity. Those workers who do not adapt to changes in technology will experience a decrease in demand.
Number of Companies	An increase in the number of companies producing a given product will increase the demand for labor resulting in a shift to the right. A decrease in the number of companies producing a given product will decrease the demand for labor resulting in a shift to the left.
Government Regulations	Complying with government regulations can increase or decrease the demand for labor at any given wage. In the healthcare industry, government rules may require that nurses be hired to carry out certain medical procedures. This will increase the demand for nurses. Less-trained healthcare workers would be prohibited from carrying out these procedures, and the demand for these workers will shift to the left.
Price and Availability of Other Inputs	Labor is not the only input into the production process. For example, a salesperson at a call center needs a telephone and a computer terminal to enter data and record sales. If prices of other inputs fall, production will become more profitable and suppliers will demand more labor to increase production. This will cause a rightward shift in the demand curve for labor. The opposite is also true. Higher prices for other inputs lower demand for labor.

Table 4.2 Factors That Can Shift Demand

Link It Up

Click here (http://openstaxcollege.org/l/Futurework) to read more about "Trends and Challenges for Work in the 21st Century."

Shifts in Labor Supply

The supply of labor is upward-sloping and adheres to the law of supply: The higher the price, the greater the quantity supplied and the lower the price, the less quantity supplied. The supply curve models the tradeoff between supplying labor into the market or using time in leisure activities at every given price level. The higher the wage, the more labor is willing to work and forego leisure activities. **Table 4.3** lists some of the factors that will cause the supply to increase or decrease.

Factors	Results
Number of Workers	An increased number of workers will cause the supply curve to shift to the right. An increased number of workers can be due to several factors, such as immigration, increasing population, an aging population, and changing demographics. Policies that encourage immigration will increase the supply of labor, and vice versa. Population grows when birth rates exceed death rates. This eventually increases supply of labor when the former reach working age. An aging and therefore retiring population will decrease the supply of labor. Another example of changing demographics is more women working outside of the home, which increases the supply of labor.
Required Education	The more required education, the lower the supply. There is a lower supply of PhD mathematicians than of high school mathematics teachers; there is a lower supply of cardiologists than of primary care physicians; and there is a lower supply of physicians than of nurses.
Government Policies	Government policies can also affect the supply of labor for jobs. Alternatively, the government may support rules that set high qualifications for certain jobs: academic training, certificates or licenses, or experience. When these qualifications are made tougher, the number of qualified workers will decrease at any given wage. On the other hand, the government may also subsidize training or even reduce the required level of qualifications. For example, government might offer subsidies for nursing schools or nursing students. Such provisions would shift the supply curve of nurses to the right. In addition, government policies that change the relative desirability of working versus not working also affect the labor supply. These include unemployment benefits, maternity leave, child care benefits, and welfare policy. For example, child care benefits may increase the labor supply of working mothers. Long term unemployment benefits may discourage job searching for unemployed workers. All these policies must therefore be carefully designed to minimize any negative labor supply effects.

Table 4.3 Factors that Can Shift Supply

A change in salary will lead to a movement along labor demand or labor supply curves, but it will not shift those curves. However, other events like those we have outlined here will cause either the demand or the supply of labor to shift, and thus will move the labor market to a new equilibrium salary and quantity.

Technology and Wage Inequality: The Four-Step Process

Economic events can change the equilibrium salary (or wage) and quantity of labor. Consider how the wave of new information technologies, like computer and telecommunications networks, has affected low-skill and high-skill workers in the U.S. economy. From the perspective of employers who demand labor, these new technologies are often a substitute for low-skill laborers like file clerks who used to keep file cabinets full of paper records of transactions. However, the same new technologies are a complement to high-skill workers like managers, who benefit from the technological advances by having the ability to monitor more information, communicate more easily, and juggle a wider array of responsibilities. How will the new technologies affect the wages of high-skill and low-skill workers? For this question, the four-step process of analyzing how shifts in supply or demand affect a market (introduced in **Demand and Supply**) works in this way:

Step 1. What did the markets for low-skill labor and high-skill labor look like before the arrival of the new technologies? In **Figure 4.3** (a) and **Figure 4.3** (b), S_0 is the original supply curve for labor and D_0 is the original demand curve for labor in each market. In each graph, the original point of equilibrium, E_0, occurs at the price W_0 and the quantity Q_0.

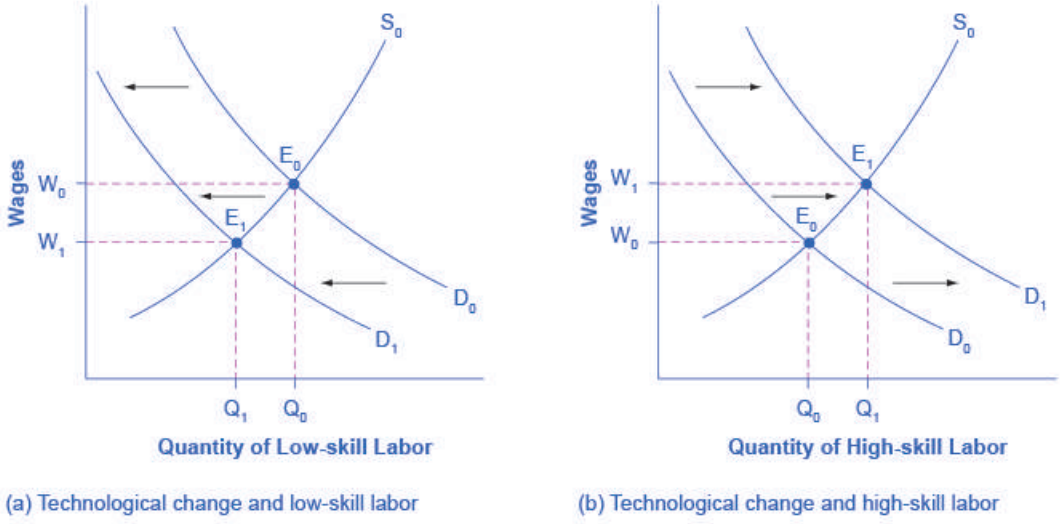


Figure 4.3 Technology and Wages: Applying Demand and Supply (a) The demand for low-skill labor shifts to the left when technology can do the job previously done by these workers. (b) New technologies can also increase the demand for high-skill labor in fields such as information technology and network administration.

Step 2. Does the new technology affect the supply of labor from households or the demand for labor from firms? The technology change described here affects demand for labor by firms that hire workers.

Step 3. Will the new technology increase or decrease demand? Based on the description earlier, as the substitute for low-skill labor becomes available, demand for low-skill labor will shift to the left, from D_0 to D_1. As the technology complement for high-skill labor becomes cheaper, demand for high-skill labor will shift to the right, from D_0 to D_1.

Step 4. The new equilibrium for low-skill labor, shown as point E_1 with price W_1 and quantity Q_1, has a lower wage and quantity hired than the original equilibrium, E_0. The new equilibrium for high-skill labor, shown as point E_1 with price W_1 and quantity Q_1, has a higher wage and quantity hired than the original equilibrium (E_0).

Thus, the demand and supply model predicts that the new computer and communications technologies will raise the pay of high-skill workers but reduce the pay of low-skill workers. From the 1970s to the mid-2000s, the wage gap widened between high-skill and low-skill labor. According to the National Center for Education Statistics, in 1980, for example, a college graduate earned about 30% more than a high school graduate with comparable job experience, but by 2014, a college graduate earned about 66% more than an otherwise comparable high school graduate. Many economists believe that the trend toward greater wage inequality across the U.S. economy that improvements in

technology.

Link It Up

Visit this website (http://openstaxcollege.org/l/oldtechjobs) to read about ten tech skills that have lost relevance in today's workforce.

Price Floors in the Labor Market: Living Wages and Minimum Wages

In contrast to goods and services markets, price ceilings are rare in labor markets, because rules that prevent people from earning income are not politically popular. There is one exception: boards of trustees or stockholders, as an example, propose limits on the high incomes of top business executives.

The labor market, however, presents some prominent examples of price floors, which are an attempt to increase the wages of low-paid workers. The U.S. government sets a **minimum wage**, a price floor that makes it illegal for an employer to pay employees less than a certain hourly rate. In mid-2009, the U.S. minimum wage was raised to $7.25 per hour. Local political movements in a number of U.S. cities have pushed for a higher minimum wage, which they call a **living wage**. Promoters of living wage laws maintain that the minimum wage is too low to ensure a reasonable standard of living. They base this conclusion on the calculation that, if you work 40 hours a week at a minimum wage of $7.25 per hour for 50 weeks a year, your annual income is $14,500, which is less than the official U.S. government definition of what it means for a family to be in poverty. (A family with two adults earning minimum wage and two young children will find it more cost efficient for one parent to provide childcare while the other works for income. Thus the family income would be $14,500, which is significantly lower than the federal poverty line for a family of four, which was $24,250 in 2015.)

Supporters of the living wage argue that full-time workers should be assured a high enough wage so that they can afford the essentials of life: food, clothing, shelter, and healthcare. Since Baltimore passed the first living wage law in 1994, several dozen cities enacted similar laws in the late 1990s and the 2000s. The living wage ordinances do not apply to all employers, but they have specified that all employees of the city or employees of firms that the city hires be paid at least a certain wage that is usually a few dollars per hour above the U.S. minimum wage.

Figure 4.4 illustrates the situation of a city considering a living wage law. For simplicity, we assume that there is no federal minimum wage. The wage appears on the vertical axis, because the wage is the price in the labor market. Before the passage of the living wage law, the equilibrium wage is $10 per hour and the city hires 1,200 workers at this wage. However, a group of concerned citizens persuades the city council to enact a living wage law requiring employers to pay no less than $12 per hour. In response to the higher wage, 1,600 workers look for jobs with the city. At this higher wage, the city, as an employer, is willing to hire only 700 workers. At the price floor, the quantity supplied exceeds the quantity demanded, and a surplus of labor exists in this market. For workers who continue to have a job at a higher salary, life has improved. For those who were willing to work at the old wage rate but lost their jobs with the wage increase, life has not improved. **Table 4.4** shows the differences in supply and demand at different wages.

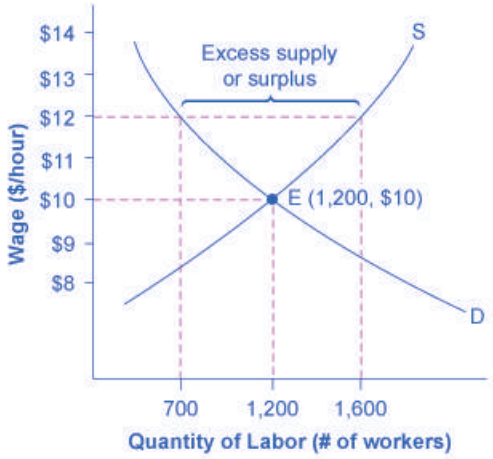


Figure 4.4 A Living Wage: Example of a Price Floor The original equilibrium in this labor market is a wage of $10/hour and a quantity of 1,200 workers, shown at point E. Imposing a wage floor at $12/hour leads to an excess supply of labor. At that wage, the quantity of labor supplied is 1,600 and the quantity of labor demanded is only 700.

Wage	Quantity Labor Demanded	Quantity Labor Supplied
$8/hr	1,900	500
$9/hr	1,500	900
$10/hr	1,200	1,200
$11/hr	900	1,400
$12/hr	700	1,600
$13/hr	500	1,800
$14/hr	400	1,900

Table 4.4 Living Wage: Example of a Price Floor

The Minimum Wage as an Example of a Price Floor

The U.S. minimum wage is a price floor that is set either very close to the equilibrium wage or even slightly below it. About 1% of American workers are actually paid the minimum wage. In other words, the vast majority of the U.S. labor force has its wages determined in the labor market, not as a result of the government price floor. However, for workers with low skills and little experience, like those without a high school diploma or teenagers, the minimum wage is quite important. In many cities, the federal minimum wage is apparently below the market price for unskilled labor, because employers offer more than the minimum wage to checkout clerks and other low-skill workers without any government prodding.

Economists have attempted to estimate how much the minimum wage reduces the quantity demanded of low-skill labor. A typical result of such studies is that a 10% increase in the minimum wage would decrease the hiring of unskilled workers by 1 to 2%, which seems a relatively small reduction. In fact, some studies have even found no effect of a higher minimum wage on employment at certain times and places—although these studies are controversial.

Let's suppose that the minimum wage lies just slightly *below* the equilibrium wage level. Wages could fluctuate according to market forces above this price floor, but they would not be allowed to move beneath the floor. In this situation, the price floor minimum wage is *nonbinding* —that is, the price floor is not determining the market outcome. Even if the minimum wage moves just a little higher, it will still have no effect on the quantity of

employment in the economy, as long as it remains below the equilibrium wage. Even if the government increases minimum wage by enough so that it rises slightly above the equilibrium wage and becomes binding, there will be only a small excess supply gap between the quantity demanded and quantity supplied.

These insights help to explain why U.S. minimum wage laws have historically had only a small impact on employment. Since the minimum wage has typically been set close to the equilibrium wage for low-skill labor and sometimes even below it, it has not had a large effect in creating an excess supply of labor. However, if the minimum wage increased dramatically—say, if it doubled to match the living wages that some U.S. cities have considered—then its impact on reducing the quantity demanded of employment would be far greater. As of 2017, many U.S. states are set to increase their minimum wage to $15 per hour. We will see what happens. The following Clear It Up feature describes in greater detail some of the arguments for and against changes to minimum wage.

Clear It Up

What's the harm in raising the minimum wage?

Because of the law of demand, a higher required wage will reduce the amount of low-skill employment either in terms of employees or in terms of work hours. Although there is controversy over the numbers, let's say for the sake of the argument that a 10% rise in the minimum wage will reduce the employment of low-skill workers by 2%. Does this outcome mean that raising the minimum wage by 10% is bad public policy? Not necessarily.

If 98% of those receiving the minimum wage have a pay increase of 10%, but 2% of those receiving the minimum wage lose their jobs, are the gains for society as a whole greater than the losses? The answer is not clear, because job losses, even for a small group, may cause more pain than modest income gains for others. For one thing, we need to consider which minimum wage workers are losing their jobs. If the 2% of minimum wage workers who lose their jobs are struggling to support families, that is one thing. If those who lose their job are high school students picking up spending money over summer vacation, that is something else.

Another complexity is that many minimum wage workers do not work full-time for an entire year. Imagine a minimum wage worker who holds different part-time jobs for a few months at a time, with bouts of unemployment in between. The worker in this situation receives the 10% raise in the minimum wage when working, but also ends up working 2% fewer hours during the year because the higher minimum wage reduces how much employers want people to work. Overall, this worker's income would rise because the 10% pay raise would more than offset the 2% fewer hours worked.

Of course, these arguments do not prove that raising the minimum wage is necessarily a good idea either. There may well be other, better public policy options for helping low-wage workers. (The **Poverty and Economic Inequality** chapter discusses some possibilities.) The lesson from this maze of minimum wage arguments is that complex social problems rarely have simple answers. Even those who agree on how a proposed economic policy affects quantity demanded and quantity supplied may still disagree on whether the policy is a good idea.

4.2 | Demand and Supply in Financial Markets

By the end of this section, you will be able to:

- Identify the demanders and suppliers in a financial market
- Explain how interest rates can affect supply and demand
- Analyze the economic effects of U.S. debt in terms of domestic financial markets
- Explain the role of price ceilings and usury laws in the U.S.

United States' households, institutions, and domestic businesses saved almost $1.3 trillion in 2015. Where did that savings go and how was it used? Some of the savings ended up in banks, which in turn loaned the money to individuals or businesses that wanted to borrow money. Some was invested in private companies or loaned to

government agencies that wanted to borrow money to raise funds for purposes like building roads or mass transit. Some firms reinvested their savings in their own businesses.

In this section, we will determine how the demand and supply model links those who wish to supply **financial capital** (i.e., savings) with those who demand financial capital (i.e., borrowing). Those who save money (or make financial investments, which is the same thing), whether individuals or businesses, are on the supply side of the financial market. Those who borrow money are on the demand side of the financial market. For a more detailed treatment of the different kinds of financial investments like bank accounts, stocks and bonds, see the **Financial Markets** chapter.

Who Demands and Who Supplies in Financial Markets?

In any market, the price is what suppliers receive and what demanders pay. In financial markets, those who supply financial capital through saving expect to receive a rate of return, while those who demand financial capital by receiving funds expect to pay a rate of return. This rate of return can come in a variety of forms, depending on the type of investment.

The simplest example of a rate of return is the **interest rate**. For example, when you supply money into a savings account at a bank, you receive interest on your deposit. The interest the bank pays you as a percent of your deposits is the interest rate. Similarly, if you demand a loan to buy a car or a computer, you will need to pay interest on the money you borrow.

Let’s consider the market for borrowing money with credit cards. In 2015, almost 200 million Americans were cardholders. Credit cards allow you to borrow money from the card's issuer, and pay back the borrowed amount plus interest, although most allow you a period of time in which you can repay the loan without paying interest. A typical credit card interest rate ranges from 12% to 18% per year. In May 2016, Americans had about $943 billion outstanding in credit card debts. About half of U.S. families with credit cards report that they almost always pay the full balance on time, but one-quarter of U.S. families with credit cards say that they “hardly ever” pay off the card in full. In fact, in 2014, 56% of consumers carried an unpaid balance in the last 12 months. Let’s say that, on average, the annual interest rate for credit card borrowing is 15% per year. Thus, Americans pay tens of billions of dollars every year in interest on their credit cards—plus basic fees for the credit card or fees for late payments.

Figure 4.5 illustrates demand and supply in the financial market for credit cards. The horizontal axis of the financial market shows the quantity of money loaned or borrowed in this market. The vertical or price axis shows the rate of return, which in the case of credit card borrowing we can measure with an interest rate. **Table 4.5** shows the quantity of financial capital that consumers demand at various interest rates and the quantity that credit card firms (often banks) are willing to supply.

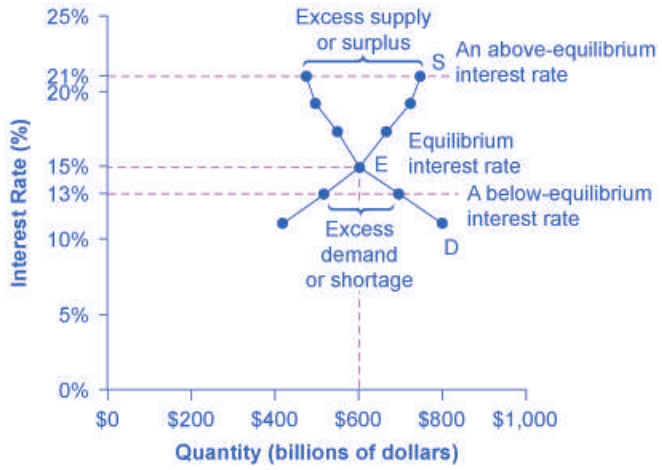


Figure 4.5 Demand and Supply for Borrowing Money with Credit Cards In this market for credit card borrowing, the demand curve (D) for borrowing financial capital intersects the supply curve (S) for lending financial capital at equilibrium E. At the equilibrium, the interest rate (the "price" in this market) is 15% and the quantity of financial capital loaned and borrowed is $600 billion. The equilibrium price is where the quantity demanded and the quantity supplied are equal. At an above-equilibrium interest rate like 21%, the quantity of financial capital supplied would increase to $750 billion, but the quantity demanded would decrease to $480 billion. At a below-equilibrium interest rate like 13%, the quantity of financial capital demanded would increase to $700 billion, but the quantity of financial capital supplied would decrease to $510 billion.

Interest Rate (%)	Quantity of Financial Capital Demanded (Borrowing) ($ billions)	Quantity of Financial Capital Supplied (Lending) ($ billions)
11	$800	$420
13	$700	$510
15	$600	$600
17	$550	$660
19	$500	$720
21	$480	$750

Table 4.5 Demand and Supply for Borrowing Money with Credit Cards

The laws of demand and supply continue to apply in the financial markets. According to the **law of demand**, a higher rate of return (that is, a higher price) will decrease the quantity demanded. As the interest rate rises, consumers will reduce the quantity that they borrow. According to the law of supply, a higher price increases the quantity supplied. Consequently, as the interest rate paid on credit card borrowing rises, more firms will be eager to issue credit cards and to encourage customers to use them. Conversely, if the interest rate on credit cards falls, the quantity of financial capital supplied in the credit card market will decrease and the quantity demanded will fall.

Equilibrium in Financial Markets

In the financial market for credit cards in **Figure 4.5**, the supply curve (S) and the demand curve (D) cross at the equilibrium point (E). The equilibrium occurs at an interest rate of 15%, where the quantity of funds demanded and the quantity supplied are equal at an equilibrium quantity of $600 billion.

If the interest rate (remember, this measures the "price" in the financial market) is above the equilibrium level, then an excess supply, or a surplus, of financial capital will arise in this market. For example, at an interest rate of 21%, the quantity of funds supplied increases to $750 billion, while the quantity demanded decreases to $480 billion. At this

above-equilibrium interest rate, firms are eager to supply loans to credit card borrowers, but relatively few people or businesses wish to borrow. As a result, some credit card firms will lower the interest rates (or other fees) they charge to attract more business. This strategy will push the interest rate down toward the equilibrium level.

If the interest rate is below the equilibrium, then excess demand or a shortage of funds occurs in this market. At an interest rate of 13%, the quantity of funds credit card borrowers demand increases to $700 billion, but the quantity credit card firms are willing to supply is only $510 billion. In this situation, credit card firms will perceive that they are overloaded with eager borrowers and conclude that they have an opportunity to raise interest rates or fees. The interest rate will face economic pressures to creep up toward the equilibrium level.

The FRED database publishes some two dozen measures of interest rates, including interest rates on credit cards, automobile loans, personal loans, mortgage loans, and more. You can find these at the FRED **website (https://openstax.org/l/FRED_stlouis)** .

Shifts in Demand and Supply in Financial Markets

Those who supply financial capital face two broad decisions: how much to save, and how to divide up their savings among different forms of financial investments. We will discuss each of these in turn.

Participants in financial markets must decide when they prefer to consume goods: now or in the future. Economists call this **intertemporal decision making** because it involves decisions across time. Unlike a decision about what to buy from the grocery store, people make investment or savings decisions across a period of time, sometimes a long period.

Most workers save for retirement because their income in the present is greater than their needs, while the opposite will be true once they retire. Thus, they save today and supply financial markets. If their income increases, they save more. If their perceived situation in the future changes, they change the amount of their saving. For example, there is some evidence that Social Security, the program that workers pay into in order to qualify for government checks after retirement, has tended to reduce the quantity of financial capital that workers save. If this is true, Social Security has shifted the supply of financial capital at any interest rate to the left.

By contrast, many college students need money today when their income is low (or nonexistent) to pay their college expenses. As a result, they borrow today and demand from financial markets. Once they graduate and become employed, they will pay back the loans. Individuals borrow money to purchase homes or cars. A business seeks financial investment so that it has the funds to build a factory or invest in a research and development project that will not pay off for five years, ten years, or even more. Thus, when consumers and businesses have greater confidence that they will be able to repay in the future, the quantity demanded of financial capital at any given interest rate will shift to the right.

For example, in the technology boom of the late 1990s, many businesses became extremely confident that investments in new technology would have a high rate of return, and their demand for financial capital shifted to the right. Conversely, during the 2008 and 2009 Great Recession, their demand for financial capital at any given interest rate shifted to the left.

To this point, we have been looking at saving in total. Now let us consider what affects saving in different types of financial investments. In deciding between different forms of financial investments, suppliers of financial capital will have to consider the rates of return and the risks involved. Rate of return is a positive attribute of investments, but risk is a negative. If Investment A becomes more risky, or the return diminishes, then savers will shift their funds to Investment B—and the supply curve of financial capital for Investment A will shift back to the left while the supply curve of capital for Investment B shifts to the right.

The United States as a Global Borrower

In the global economy, trillions of dollars of financial investment cross national borders every year. In the early 2000s, financial investors from foreign countries were investing several hundred billion dollars per year more in the U.S. economy than U.S. financial investors were investing abroad. The following Work It Out deals with one of the macroeconomic concerns for the U.S. economy in recent years.

Work It Out

The Effect of Growing U.S. Debt

Imagine that foreign investors viewed the U.S. economy as a less desirable place to put their money because of fears about the growth of the U.S. public debt. Using the four-step process for analyzing how changes in supply and demand affect equilibrium outcomes, how would increased U.S. public debt affect the equilibrium price and quantity for capital in U.S. financial markets?

Step 1. Draw a diagram showing demand and supply for financial capital that represents the original scenario in which foreign investors are pouring money into the U.S. economy. **Figure 4.6** shows a demand curve, D, and a supply curve, S, where the supply of capital includes the funds arriving from foreign investors. The original equilibrium E_0 occurs at interest rate R_0 and quantity of financial investment Q_0.

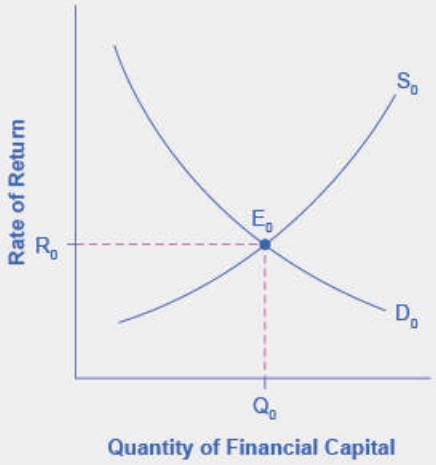


Figure 4.6 The United States as a Global Borrower Before U.S. Debt Uncertainty The graph shows the demand for financial capital from and supply of financial capital into the U.S. financial markets by the foreign sector before the increase in uncertainty regarding U.S. public debt. The original equilibrium (E_0) occurs at an equilibrium rate of return (R_0) and the equilibrium quantity is at Q_0.

Step 2. Will the diminished confidence in the U.S. economy as a place to invest affect demand or supply of financial capital? Yes, it will affect supply. Many foreign investors look to the U.S. financial markets to store their money in safe financial vehicles with low risk and stable returns. Diminished confidence means U.S. financial assets will be seen as more risky.

Step 3. Will supply increase or decrease? When the enthusiasm of foreign investors' for investing their money in the U.S. economy diminishes, the supply of financial capital shifts to the left. **Figure 4.7** shows the supply curve shift from S_0 to S_1.

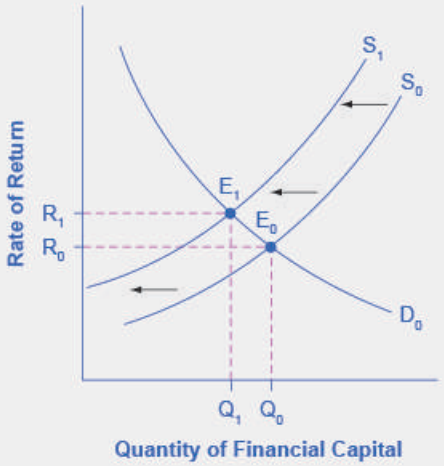


Figure 4.7 The United States as a Global Borrower Before and After U.S. Debt Uncertainty The graph shows the demand for financial capital and supply of financial capital into the U.S. financial markets by the foreign sector before and after the increase in uncertainty regarding U.S. public debt. The original equilibrium (E_0) occurs at an equilibrium rate of return (R_0) and the equilibrium quantity is at Q_0.

Step 4. Thus, foreign investors' diminished enthusiasm leads to a new equilibrium, E_1, which occurs at the higher interest rate, R_1, and the lower quantity of financial investment, Q_1. In short, U.S. borrowers will have to pay more interest on their borrowing.

The economy has experienced an enormous inflow of foreign capital. According to the U.S. Bureau of Economic Analysis, by the third quarter of 2015, U.S. investors had accumulated $23.3 trillion of foreign assets, but foreign investors owned a total of $30.6 trillion of U.S. assets. If foreign investors were to pull their money out of the U.S. economy and invest elsewhere in the world, the result could be a significantly lower quantity of financial investment in the United States, available only at a higher interest rate. This reduced inflow of foreign financial investment could impose hardship on U.S. consumers and firms interested in borrowing.

In a modern, developed economy, financial capital often moves invisibly through electronic transfers between one bank account and another. Yet we can analyze these flows of funds with the same tools of demand and supply as markets for goods or labor.

Price Ceilings in Financial Markets: Usury Laws

As we noted earlier, about 200 million Americans own credit cards, and their interest payments and fees total tens of billions of dollars each year. It is little wonder that political pressures sometimes arise for setting limits on the interest rates or fees that credit card companies charge. The firms that issue credit cards, including banks, oil companies, phone companies, and retail stores, respond that the higher interest rates are necessary to cover the losses created by those who borrow on their credit cards and who do not repay on time or at all. These companies also point out that cardholders can avoid paying interest if they pay their bills on time.

Consider the credit card market as **Figure 4.8** illustrators. In this financial market, the vertical axis shows the interest rate (which is the price in the financial market). Demanders in the credit card market are households and businesses. Suppliers are the companies that issue credit cards. This figure does not use specific numbers, which would be hypothetical in any case, but instead focuses on the underlying economic relationships. Imagine a law imposes a price ceiling that holds the interest rate charged on credit cards at the rate Rc, which lies below the interest rate R_0 that would otherwise have prevailed in the market. The horizontal dashed line at interest rate R_c in **Figure 4.8** shows the price ceiling. The demand and supply model predicts that at the lower price ceiling interest rate, the quantity demanded of credit card debt will increase from its original level of Q_0 to Qd; however, the quantity supplied of credit card debt will decrease from the original Q_0 to Qs. At the price ceiling (Rc), quantity demanded will exceed quantity supplied. Consequently, a number of people who want to have credit cards and are willing to pay the prevailing interest rate will find that companies are unwilling to issue cards to them. The result will be a credit shortage.

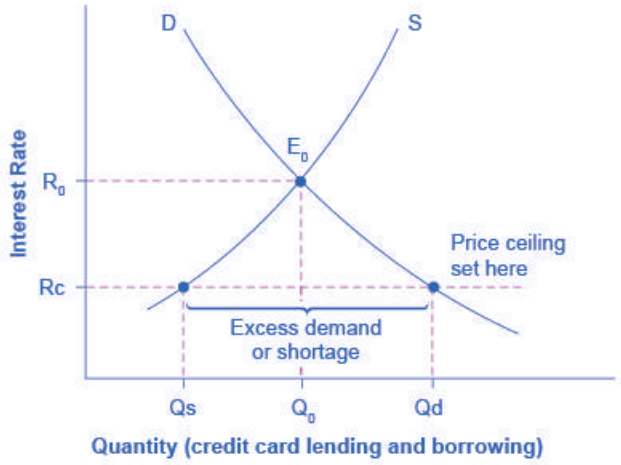


Figure 4.8 Credit Card Interest Rates: Another Price Ceiling Example The original intersection of demand D and supply S occurs at equilibrium E_0. However, a price ceiling is set at the interest rate Rc, below the equilibrium interest rate R_0, and so the interest rate cannot adjust upward to the equilibrium. At the price ceiling, the quantity demanded, Qd, exceeds the quantity supplied, Qs. There is excess demand, also called a shortage.

Many states do have **usury laws**, which impose an upper limit on the interest rate that lenders can charge. However, in many cases these upper limits are well above the market interest rate. For example, if the interest rate is not allowed to rise above 30% per year, it can still fluctuate below that level according to market forces. A price ceiling that is set at a relatively high level is nonbinding, and it will have no practical effect unless the equilibrium price soars high enough to exceed the price ceiling.

4.3 | The Market System as an Efficient Mechanism for Information

By the end of this section, you will be able to:

- Apply demand and supply models to analyze prices and quantities
- Explain the effects of price controls on the equilibrium of prices and quantities

Prices exist in markets for goods and services, for labor, and for financial capital. In all of these markets, prices serve as a remarkable social mechanism for collecting, combining, and transmitting information that is relevant to the market—namely, the relationship between demand and supply—and then serving as messengers to convey that information to buyers and sellers. In a market-oriented economy, no government agency or guiding intelligence oversees the set of responses and interconnections that result from a change in price. Instead, each consumer reacts according to that person's preferences and budget set, and each profit-seeking producer reacts to the impact on its expected profits. The following Clear It Up feature examines the **demand and supply models**.

Clear It Up

Why are demand and supply curves important?

The demand and supply model is the second fundamental diagram for this course. (The opportunity set model that we introduced in the **Choice in a World of Scarcity** chapter was the first.) Just as it would be foolish to try to learn the arithmetic of long division by memorizing every possible combination of numbers that can be divided by each other, it would be foolish to try to memorize every specific example of demand and supply in this chapter, this textbook, or this course. Demand and supply is not primarily a list of examples. It is a

model to analyze prices and quantities. Even though demand and supply diagrams have many labels, they are fundamentally the same in their logic. Your goal should be to understand the underlying model so you can use it to analyze *any* market.

Figure 4.9 displays a generic demand and supply curve. The horizontal axis shows the different measures of quantity: a quantity of a good or service, or a quantity of labor for a given job, or a quantity of financial capital. The vertical axis shows a measure of price: the price of a good or service, the wage in the labor market, or the rate of return (like the interest rate) in the financial market.

The demand and supply model can explain the existing levels of prices, wages, and rates of return. To carry out such an analysis, think about the quantity that will be demanded at each price and the quantity that will be supplied at each price—that is, think about the shape of the demand and supply curves—and how these forces will combine to produce equilibrium.

We can also use demand and supply to explain how economic events will cause changes in prices, wages, and rates of return. There are only four possibilities: the change in any single event may cause the demand curve to shift right or to shift left, or it may cause the supply curve to shift right or to shift left. The key to analyzing the effect of an economic event on equilibrium prices and quantities is to determine which of these four possibilities occurred. The way to do this correctly is to think back to the list of factors that shift the demand and supply curves. Note that if more than one variable is changing at the same time, the overall impact will depend on the degree of the shifts. When there are multiple variables, economists isolate each change and analyze it independently.

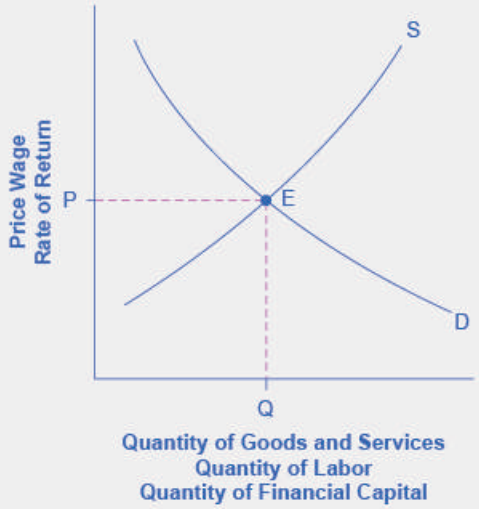


Figure 4.9 Demand and Supply Curves The figure displays a generic demand and supply curve. The horizontal axis shows the different measures of quantity: a quantity of a good or service, a quantity of labor for a given job, or a quantity of financial capital. The vertical axis shows a measure of price: the price of a good or service, the wage in the labor market, or the rate of return (like the interest rate) in the financial market. We can use the demand and supply curves explain how economic events will cause changes in prices, wages, and rates of return.

An increase in the price of some product signals consumers that there is a shortage; therefore, they may want to economize on buying this product. For example, if you are thinking about taking a plane trip to Hawaii, but the ticket turns out to be expensive during the week you intend to go, you might consider other weeks when the ticket might be cheaper. The price could be high because you were planning to travel during a holiday when demand for traveling is high. Maybe the cost of an input like jet fuel increased or the airline has raised the price temporarily to see how many people are willing to pay it. Perhaps all of these factors are present at the same time. You do not need to analyze the market and break down the price change into its underlying factors. You just have to look at the ticket price and decide whether and when to fly.

In the same way, price changes provide useful information to producers. Imagine the situation of a farmer who grows

oats and learns that the price of oats has risen. The higher price could be due to an increase in demand caused by a new scientific study proclaiming that eating oats is especially healthful. Perhaps the price of a substitute grain, like corn, has risen, and people have responded by buying more oats. The oat farmer does not need to know the details. The farmer only needs to know that the price of oats has risen and that it will be profitable to expand production as a result.

The actions of individual consumers and producers as they react to prices overlap and interlock in markets for goods, labor, and financial capital. A change in any single market is transmitted through these multiple interconnections to other markets. The vision of the role of flexible prices helping markets to reach equilibrium and linking different markets together helps to explain why price controls can be so counterproductive. Price controls are government laws that serve to regulate prices rather than allow the various markets to determine prices. There is an old proverb: "Don't kill the messenger." In ancient times, messengers carried information between distant cities and kingdoms. When they brought bad news, there was an emotional impulse to kill the messenger. However, killing the messenger did not kill the bad news. Moreover, killing the messenger had an undesirable side effect: Other messengers would refuse to bring news to that city or kingdom, depriving its citizens of vital information.

Those who seek price controls are trying to kill the messenger—or at least to stifle an unwelcome message that prices are bringing about the equilibrium level of price and quantity. However, price controls do nothing to affect the underlying forces of demand and supply, and this can have serious repercussions. During China's "Great Leap Forward" in the late 1950s, the government kept food prices artificially low, with the result that 30 to 40 million people died of starvation because the low prices depressed farm production. This was communist party leader Mao Zedong's social and economic campaign to rapidly transform the country from an agrarian economy to a socialist society through rapid industrialization and collectivization. Changes in demand and supply will continue to reveal themselves through consumers' and producers' behavior. Immobilizing the price messenger through price controls will deprive everyone in the economy of critical information. Without this information, it becomes difficult for everyone—buyers and sellers alike—to react in a flexible and appropriate manner as changes occur throughout the economy.

Bring it Home

Baby Boomers Come of Age

The theory of supply and demand can explain what happens in the labor markets and suggests that the demand for nurses will increase as healthcare needs of baby boomers increase, as **Figure 4.10** shows. The impact of that increase will result in an average salary higher than the $67,490 earned in 2015 referenced in the first part of this case. The new equilibrium (E_1) will be at the new equilibrium price (Pe_1).Equilibrium quantity will also increase from Qe_0 to Qe_1.

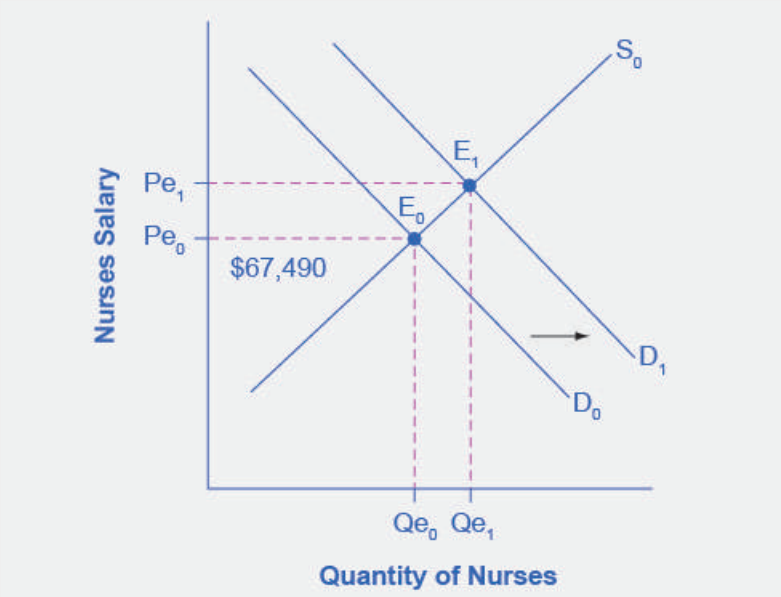


Figure 4.10 Impact of Increasing Demand for Nurses 2014-2024 In 2014, the median salary for nurses was $67,490. As demand for services increases, the demand curve shifts to the right (from D_0 to D_1) and the equilibrium quantity of nurses increases from Qe_0 to Qe_1. The equilibrium salary increases from Pe_0 to Pe_1.

Suppose that as the demand for nurses increases, the supply shrinks due to an increasing number of nurses entering retirement and increases in the tuition of nursing degrees. The leftward shift of the supply curve in **Figure 4.11** captures the impact of a decreasing supply of nurses. The shifts in the two curves result in higher salaries for nurses, but the overall impact in the quantity of nurses is uncertain, as it depends on the relative shifts of supply and demand.

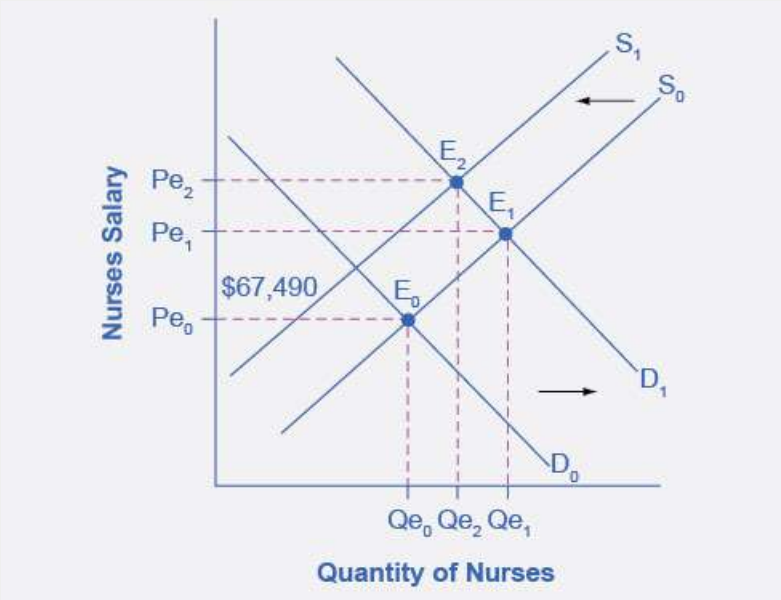


Figure 4.11 Impact of Decreasing Supply of Nurses between 2014 and 2024 The increase in demand for nurses shown in Figure 4.10 leads to both higher prices and higher quantities demanded. As nurses retire from the work force, the supply of nurses decreases, causing a leftward shift in the supply curve and higher salaries for nurses at Pe_2. The net effect on the equilibrium quantity of nurses is uncertain, which in this representation is less than Qe_1, but more than the initial Qe_0.

While we do not know if the number of nurses will increase or decrease relative to their initial employment, we know they will have higher salaries.

KEY TERMS

interest rate the "price" of borrowing in the financial market; a rate of return on an investment

minimum wage a price floor that makes it illegal for an employer to pay employees less than a certain hourly rate

usury laws laws that impose an upper limit on the interest rate that lenders can charge

KEY CONCEPTS AND SUMMARY

4.1 Demand and Supply at Work in Labor Markets

In the labor market, households are on the supply side of the market and firms are on the demand side. In the market for financial capital, households and firms can be on either side of the market: they are suppliers of financial capital when they save or make financial investments, and demanders of financial capital when they borrow or receive financial investments.

In the demand and supply analysis of labor markets, we can measure the price by the annual salary or hourly wage received. We can measure the quantity of labor various ways, like number of workers or the number of hours worked.

Factors that can shift the demand curve for labor include: a change in the quantity demanded of the product that the labor produces; a change in the production process that uses more or less labor; and a change in government policy that affects the quantity of labor that firms wish to hire at a given wage. Demand can also increase or decrease (shift) in response to: workers' level of education and training, technology, the number of companies, and availability and price of other inputs.

The main factors that can shift the supply curve for labor are: how desirable a job appears to workers relative to the alternatives, government policy that either restricts or encourages the quantity of workers trained for the job, the number of workers in the economy, and required education.

4.2 Demand and Supply in Financial Markets

In the demand and supply analysis of financial markets, the "price" is the rate of return or the interest rate received. We measure the quantity by the money that flows from those who supply financial capital to those who demand it.

Two factors can shift the supply of financial capital to a certain investment: if people want to alter their existing levels of consumption, and if the riskiness or return on one investment changes relative to other investments. Factors that can shift demand for capital include business confidence and consumer confidence in the future—since financial investments received in the present are typically repaid in the future.

4.3 The Market System as an Efficient Mechanism for Information

The market price system provides a highly efficient mechanism for disseminating information about relative scarcities of goods, services, labor, and financial capital. Market participants do not need to know why prices have changed, only that the changes require them to revisit previous decisions they made about supply and demand. Price controls hide information about the true scarcity of products and thereby cause misallocation of resources.

SELF-CHECK QUESTIONS

1. In the labor market, what causes a movement along the demand curve? What causes a shift in the demand curve?

2. In the labor market, what causes a movement along the supply curve? What causes a shift in the supply curve?

3. Why is a living wage considered a price floor? Does imposing a living wage have the same outcome as a minimum wage?

4. In the financial market, what causes a movement along the demand curve? What causes a shift in the demand curve?

5. In the financial market, what causes a movement along the supply curve? What causes a shift in the supply curve?

6. If a usury law limits interest rates to no more than 35%, what would the likely impact be on the amount of loans made and interest rates paid?

7. Which of the following changes in the financial market will lead to a decline in interest rates:
a. a rise in demand
b. a fall in demand
c. a rise in supply
d. a fall in supply

8. Which of the following changes in the financial market will lead to an increase in the quantity of loans made and received:
a. a rise in demand
b. a fall in demand
c. a rise in supply
d. a fall in supply

9. Identify the most accurate statement. A price floor will have the largest effect if it is set:
a. substantially above the equilibrium price
b. slightly above the equilibrium price
c. slightly below the equilibrium price
d. substantially below the equilibrium price

Sketch all four of these possibilities on a demand and supply diagram to illustrate your answer.

10. A price ceiling will have the largest effect:
a. substantially below the equilibrium price
b. slightly below the equilibrium price
c. substantially above the equilibrium price
d. slightly above the equilibrium price

Sketch all four of these possibilities on a demand and supply diagram to illustrate your answer.

11. Select the correct answer. A price floor will usually shift:
a. demand
b. supply
c. both
d. neither

Illustrate your answer with a diagram.

12. Select the correct answer. A price ceiling will usually shift:
a. demand
b. supply
c. both
d. neither

REVIEW QUESTIONS

13. What is the “price” commonly called in the labor market?

14. Are households demanders or suppliers in the goods market? Are firms demanders or suppliers in the goods market? What about the labor market and the financial market?

15. Name some factors that can cause a shift in the demand curve in labor markets.

16. Name some factors that can cause a shift in the supply curve in labor markets.

17. How do economists define equilibrium in financial markets?

18. What would be a sign of a shortage in financial markets?

19. Would usury laws help or hinder resolution of a shortage in financial markets?

20. Whether the product market or the labor market, what happens to the equilibrium price and quantity for each of the four possibilities: increase in demand, decrease in demand, increase in supply, and decrease in supply.

CRITICAL THINKING QUESTIONS

21. Other than the demand for labor, what would be another example of a "derived demand?"

22. Suppose that a 5% increase in the minimum wage causes a 5% reduction in employment. How would this affect employers and how would it affect workers? In your opinion, would this be a good policy?

23. Under what circumstances would a minimum wage be a nonbinding price floor? Under what circumstances would a living raise be a binding price floor?

24. Suppose the U.S. economy began to grow more rapidly than other countries in the world. What would be the likely impact on U.S. financial markets as part of the global economy?

25. If the government imposed a federal interest rate ceiling of 20% on all loans, who would gain and who would lose?

26. Why are the factors that shift the demand for a product different from the factors that shift the demand for labor? Why are the factors that shift the supply of a product different from those that shift the supply of labor?

27. During a discussion several years ago on building a pipeline to Alaska to carry natural gas, the U.S. Senate passed a bill stipulating that there should be a guaranteed minimum price for the natural gas that would flow through the pipeline. The thinking behind the bill was that if private firms had a guaranteed price for their natural gas, they would be more willing to drill for gas and to pay to build the pipeline.

a. Using the demand and supply framework, predict the effects of this price floor on the price, quantity demanded, and quantity supplied.
b. With the enactment of this price floor for natural gas, what are some of the likely unintended consequences in the market?
c. Suggest some policies other than the price floor that the government can pursue if it wishes to encourage drilling for natural gas and for a new pipeline in Alaska.

PROBLEMS

28. Identify each of the following as involving either demand or supply. Draw a circular flow diagram and label the flows A through F. (Some choices can be on both sides of the goods market.)

a. Households in the labor market
b. Firms in the goods market
c. Firms in the financial market
d. Households in the goods market
e. Firms in the labor market
f. Households in the financial market

29. Predict how each of the following events will raise or lower the equilibrium wage and quantity of oil workers in Texas. In each case, sketch a demand and supply diagram to illustrate your answer.

a. The price of coal rises.
b. New oil-drilling equipment is invented that is cheap and requires few workers to run.
c. Several major companies that do not mine coal open factories in Texas, offering many well-paid jobs outside the oil industry.
d. Government imposes costly new regulations to make oil-drilling a safer job.

30. Predict how each of the following economic changes will affect the equilibrium price and quantity in the financial market for home loans. Sketch a demand and supply diagram to support your answers.

a. The number of people at the most common ages for home-buying increases.
b. People gain confidence that the economy is growing and that their jobs are secure.
c. Banks that have made home loans find that a larger number of people than they expected are not repaying those loans.
d. Because of a threat of a war, people become uncertain about their economic future.
e. The overall level of saving in the economy diminishes.
f. The federal government changes its bank regulations in a way that makes it cheaper and easier for banks to make home loans.

31. Table 4.6 shows the amount of savings and borrowing in a market for loans to purchase homes, measured in millions of dollars, at various interest rates. What is the equilibrium interest rate and quantity in the capital financial market? How can you tell? Now, imagine that because of a shift in the perceptions of foreign investors, the supply curve shifts so that there will be $10 million less supplied at every interest rate. Calculate the new equilibrium interest rate and quantity, and explain why the direction of the interest rate shift makes intuitive sense.

Interest Rate	Qs	Qd
5%	130	170
6%	135	150
7%	140	140
8%	145	135
9%	150	125
10%	155	110

Table 4.6

32. Imagine that to preserve the traditional way of life in small fishing villages, a government decides to impose a price floor that will guarantee all fishermen a certain price for their catch.

a. Using the demand and supply framework, predict the effects on the price, quantity demanded, and quantity supplied.
b. With the enactment of this price floor for fish, what are some of the likely unintended consequences in the market?
c. Suggest some policies other than the price floor to make it possible for small fishing villages to continue.

33. What happens to the price and the quantity bought and sold in the cocoa market if countries producing cocoa experience a drought and a new study is released demonstrating the health benefits of cocoa? Illustrate your answer with a demand and supply graph.

5 | Elasticity

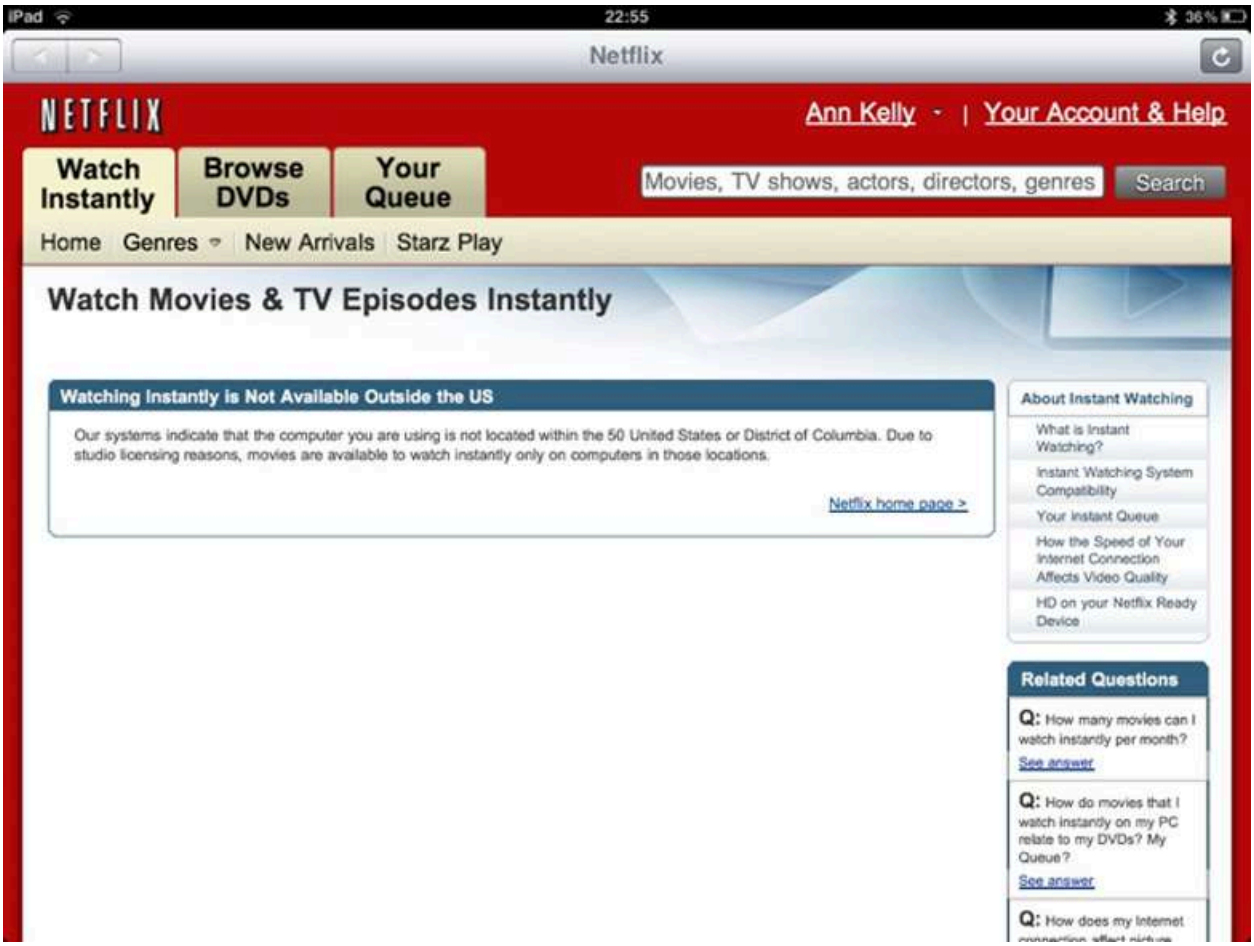


Figure 5.1 Netflix On-Demand Media Netflix, Inc. is an American provider of on-demand Internet streaming media to many countries around the world, including the United States, and of flat rate DVD-by-mail in the United States. (Credit: modification of work by Traci Lawson/Flickr Creative Commons)

Bring it Home

That Will Be How Much?

Imagine going to your favorite coffee shop and having the waiter inform you the pricing has changed. Instead of $3 for a cup of coffee, you will now be charged $2 for coffee, $1 for creamer, and $1 for your choice of sweetener. If you pay your usual $3 for a cup of coffee, you must choose between creamer and sweetener. If you want both, you now face an extra charge of $1. Sound absurd? Well, that is similar to the situation Netflix customers found themselves in—they faced a 60% price hike to retain the same service in 2011.

In early 2011, Netflix consumers paid about $10 a month for a package consisting of streaming video and DVD rentals. In July 2011, the company announced a packaging change. Customers wishing to retain both streaming video and DVD rental would be charged $15.98 per month, a price increase of about 60%. In 2014, Netflix also raised its streaming video subscription price from $7.99 to $8.99 per month for new U.S. customers. The company also changed its policy of 4K streaming content from $9.00 to $12.00 per month that year.

How would customers of the 18-year-old firm react? Would they abandon Netflix? Would the ease of access to other venues make a difference in how consumers responded to the Netflix price change? We will explore the answers to those questions in this chapter, which focuses on the change in quantity with respect to a change in price, a concept economists call elasticity.

Introduction to Elasticity

In this chapter, you will learn about:

- Price Elasticity of Demand and Price Elasticity of Supply
- Polar Cases of Elasticity and Constant Elasticity
- Elasticity and Pricing
- Elasticity in Areas Other Than Price

Anyone who has studied economics knows the law of demand: a higher price will lead to a lower quantity demanded. What you may not know is how much lower the quantity demanded will be. Similarly, the law of supply states that a higher price will lead to a higher quantity supplied. The question is: How much higher? This chapter will explain how to answer these questions and why they are critically important in the real world.

To find answers to these questions, we need to understand the concept of elasticity. **Elasticity** is an economics concept that measures responsiveness of one variable to changes in another variable. Suppose you drop two items from a second-floor balcony. The first item is a tennis ball. The second item is a brick. Which will bounce higher? Obviously, the tennis ball. We would say that the tennis ball has greater elasticity.

Consider an economic example. Cigarette taxes are an example of a "sin tax," a tax on something that is bad for you, like alcohol. Governments tax cigarettes at the state and national levels. State taxes range from a low of 17 cents per pack in Missouri to $4.35 per pack in New York. The average state cigarette tax is $1.69 per pack. The 2014 federal tax rate on cigarettes was $1.01 per pack, but in 2015 the Obama Administration proposed raising the federal tax nearly a dollar to $1.95 per pack. The key question is: How much would cigarette purchases decline?

Taxes on cigarettes serve two purposes: to raise tax revenue for government and to discourage cigarette consumption. However, if a higher cigarette tax discourages consumption considerably, meaning a greatly reduced quantity of cigarette sales, then the cigarette tax on each pack will not raise much revenue for the government. Alternatively, a higher cigarette tax that does not discourage consumption by much will actually raise more tax revenue for the government. Thus, when a government agency tries to calculate the effects of altering its cigarette tax, it must analyze how much the tax affects the quantity of cigarettes consumed. This issue reaches beyond governments and taxes. Every firm faces a similar issue. When a firm considers raising the sales price, it must consider how much a price increase will reduce the quantity demanded of what it sells. Conversely, when a firm puts its products on sale, it must expect (or hope) that the lower price will lead to a significantly higher quantity demanded.

5.1 | Price Elasticity of Demand and Price Elasticity of Supply

By the end of this section, you will be able to:

- Calculate the price elasticity of demand
- Calculate the price elasticity of supply

Both the demand and supply curve show the relationship between price and the number of units demanded or supplied. **Price elasticity** is the ratio between the percentage change in the quantity demanded (Qd) or supplied (Qs) and the corresponding percent change in price. The **price elasticity of demand** is the percentage change in the quantity *demanded* of a good or service divided by the percentage change in the price. The **price elasticity of supply** is the percentage change in quantity *supplied* divided by the percentage change in price.

We can usefully divide elasticities into three broad categories: elastic, inelastic, and unitary. An **elastic demand** or **elastic supply** is one in which the elasticity is greater than one, indicating a high responsiveness to changes in price. Elasticities that are less than one indicate low responsiveness to price changes and correspond to **inelastic demand** or **inelastic supply**. **Unitary elasticities** indicate proportional responsiveness of either demand or supply, as **Table 5.1** summarizes.

If . . .	Then . . .	And It Is Called . . .
% change in quantity > % change in price	$\frac{\%\text{ change in quantity}}{\%\text{ change in price}} > 1$	Elastic
% change in quantity = % change in price	$\frac{\%\text{ change in quantity}}{\%\text{ change in price}} = 1$	Unitary
% change in quantity < % change in price	$\frac{\%\text{ change in quantity}}{\%\text{ change in price}} < 1$	Inelastic

Table 5.1 Elastic, Inelastic, and Unitary: Three Cases of Elasticity

Link It Up

Before we delve into the details of elasticity, enjoy this article (http://openstaxcollege.org/l/Super_Bowl) on elasticity and ticket prices at the Super Bowl.

To calculate elasticity along a demand or supply curve economists use the average percent change in both quantity and price. This is called the Midpoint Method for Elasticity, and is represented in the following equations:

$$\%\text{ change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1)/2} \times 100$$

$$\%\text{ change in price} = \frac{P_2 - P_1}{(P_2 + P_1)/2} \times 100$$

The advantage of the Midpoint Method is that one obtains the same elasticity between two price points whether there is a price increase or decrease. This is because the formula uses the same base (average quantity and average price) for both cases.

Calculating Price Elasticity of Demand

Let's calculate the elasticity between points A and B and between points G and H as **Figure 5.2** shows.

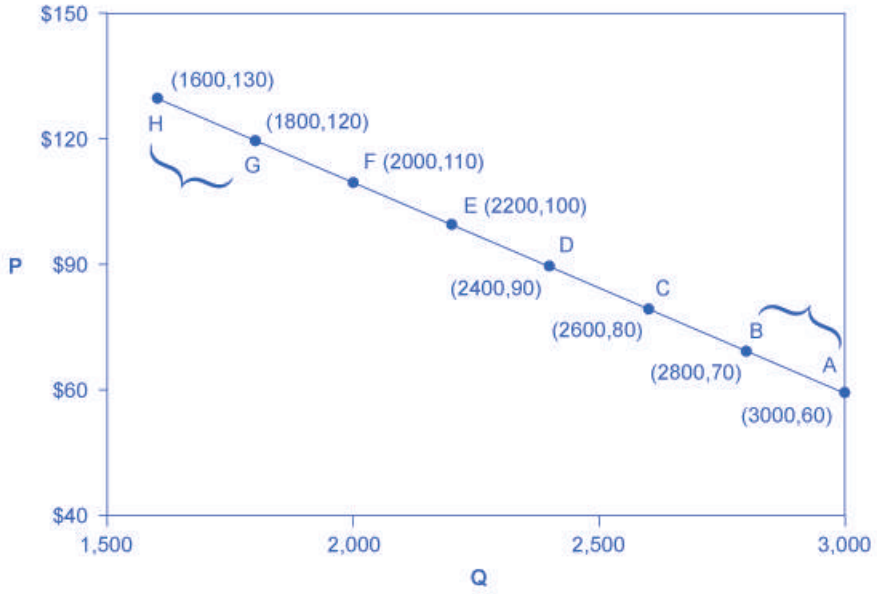


Figure 5.2 Calculating the Price Elasticity of Demand We calculate the price elasticity of demand as the percentage change in quantity divided by the percentage change in price.

First, apply the formula to calculate the elasticity as price decreases from \$70 at point B to \$60 at point A:

$$\begin{aligned}
\%\text{ change in quantity} &= \frac{3{,}000 - 2{,}800}{(3{,}000 + 2{,}800)/2} \times 100 \\
&= \frac{200}{2{,}900} \times 100 \\
&= 6.9 \\
\%\text{ change in price} &= \frac{60 - 70}{(60 + 70)/2} \times 100 \\
&= \frac{-10}{65} \times 100 \\
&= -15.4 \\
\text{Price Elasticity of Demand} &= \frac{6.9\%}{-15.4\%} \\
&= 0.45
\end{aligned}$$

Therefore, the elasticity of demand between these two points is $\frac{6.9\%}{-15.4\%}$ which is 0.45, an amount smaller than one, showing that the demand is inelastic in this interval. Price elasticities of demand are *always* negative since price and quantity demanded always move in opposite directions (on the demand curve). By convention, we always talk about elasticities as positive numbers. Mathematically, we take the absolute value of the result. We will ignore this detail from now on, while remembering to interpret elasticities as positive numbers.

This means that, along the demand curve between point B and A, if the price changes by 1%, the quantity demanded will change by 0.45%. A change in the price will result in a smaller percentage change in the quantity demanded. For example, a 10% *increase* in the price will result in only a 4.5% *decrease* in quantity demanded. A 10% *decrease* in the price will result in only a 4.5% *increase* in the quantity demanded. Price elasticities of demand are negative numbers indicating that the demand curve is downward sloping, but we read them as absolute values. The following Work It Out feature will walk you through calculating the price elasticity of demand.

Work It Out

Finding the Price Elasticity of Demand

Calculate the price elasticity of demand using the data in **Figure 5.2** for an increase in price from G to H. Has the elasticity increased or decreased?

Step 1. We know that:

$$\text{Price Elasticity of Demand} = \frac{\%\text{ change in quantity}}{\%\text{ change in price}}$$

Step 2. From the Midpoint Formula we know that:

$$\%\text{ change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1)/2} \times 100$$

$$\%\text{ change in price} = \frac{P_2 - P_1}{(P_2 + P_1)/2} \times 100$$

Step 3. So we can use the values provided in the figure in each equation:

$$\begin{aligned} \%\text{ change in quantity} &= \frac{1{,}600 - 1{,}800}{(1{,}600 + 1{,}800)/2} \times 100 \\ &= \frac{-200}{1{,}700} \times 100 \\ &= -11.76 \\ \%\text{ change in price} &= \frac{130 - 120}{(130 + 120)/2} \times 100 \\ &= \frac{10}{125} \times 100 \\ &= 8.0 \end{aligned}$$

Step 4. Then, we can use those values to determine the price elasticity of demand:

$$\begin{aligned} \text{Price Elasticity of Demand} &= \frac{\%\text{ change in quantity}}{\%\text{ change in price}} \\ &= \frac{-11.76}{8} \\ &= 1.47 \end{aligned}$$

Therefore, the elasticity of demand from G to is H 1.47. The magnitude of the elasticity has increased (in absolute value) as we moved up along the demand curve from points A to B. Recall that the elasticity between these two points was 0.45. Demand was inelastic between points A and B and elastic between points G and H. This shows us that price elasticity of demand changes at different points along a straight-line demand curve.

Calculating the Price Elasticity of Supply

Assume that an apartment rents for $650 per month and at that price the landlord rents 10,000 units are rented as **Figure 5.3** shows. When the price increases to $700 per month, the landlord supplies 13,000 units into the market. By what percentage does apartment supply increase? What is the price sensitivity?

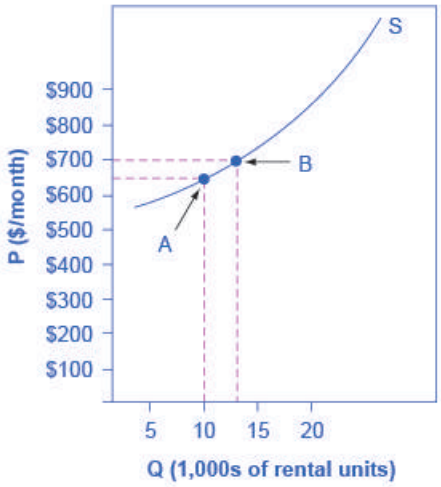


Figure 5.3 Price Elasticity of Supply We calculate the price elasticity of supply as the percentage change in quantity divided by the percentage change in price.

Using the Midpoint Method,

$$\begin{aligned}
\%\text{ change in quantity} &= \frac{13{,}000 - 10{,}000}{(13{,}000 + 10{,}000)/2} \times 100 \\
&= \frac{3{,}000}{11{,}500} \times 100 \\
&= 26.1 \\
\%\text{ change in price} &= \frac{\$700 - \$650}{(\$700 + \$650)/2} \times 100 \\
&= \frac{50}{675} \times 100 \\
&= 7.4 \\
\text{Price Elasticity of Supply} &= \frac{26.1\%}{7.4\%} \\
&= 3.53
\end{aligned}$$

Again, as with the elasticity of demand, the elasticity of supply is not followed by any units. Elasticity is a ratio of one percentage change to another percentage change—nothing more—and we read it as an absolute value. In this case, a 1% rise in price causes an increase in quantity supplied of 3.5%. The greater than one elasticity of supply means that the percentage change in quantity supplied will be greater than a one percent price change. If you're starting to wonder if the concept of slope fits into this calculation, read the following Clear It Up box.

Clear It Up

Is the elasticity the slope?

It is a common mistake to confuse the slope of either the supply or demand curve with its elasticity. The slope is the rate of change in units along the curve, or the rise/run (change in y over the change in x). For example, in **Figure 5.2**, at each point shown on the demand curve, price drops by $10 and the number of units demanded increases by 200 compared to the point to its left. The slope is –10/200 along the entire demand curve and does not change. The price elasticity, however, changes along the curve. Elasticity between points A and B was 0.45 and increased to 1.47 between points G and H. Elasticity is the *percentage* change, which is a different calculation from the slope and has a different meaning.

When we are at the upper end of a demand curve, where price is high and the quantity demanded is low, a small change in the quantity demanded, even in, say, one unit, is pretty big in percentage terms. A change

in price of, say, a dollar, is going to be much less important in percentage terms than it would have been at the bottom of the demand curve. Likewise, at the bottom of the demand curve, that one unit change when the quantity demanded is high will be small as a percentage.

Thus, at one end of the demand curve, where we have a large percentage change in quantity demanded over a small percentage change in price, the elasticity value would be high, or demand would be relatively elastic. Even with the same change in the price and the same change in the quantity demanded, at the other end of the demand curve the quantity is much higher, and the price is much lower, so the percentage change in quantity demanded is smaller and the percentage change in price is much higher. That means at the bottom of the curve we'd have a small numerator over a large denominator, so the elasticity measure would be much lower, or inelastic.

As we move along the demand curve, the values for quantity and price go up or down, depending on which way we are moving, so the percentages for, say, a $1 difference in price or a one unit difference in quantity, will change as well, which means the ratios of those percentages and hence the elasticity will change.

5.2 | Polar Cases of Elasticity and Constant Elasticity

By the end of this section, you will be able to:

- Differentiate between infinite and zero elasticity
- Analyze graphs in order to classify elasticity as constant unitary, infinite, or zero

There are two extreme cases of elasticity: when elasticity equals zero and when it is infinite. A third case is that of constant unitary elasticity. We will describe each case. **Infinite elasticity** or **perfect elasticity** refers to the extreme case where either the quantity demanded (Qd) or supplied (Qs) changes by an infinite amount in response to any change in price at all. In both cases, the supply and the demand curve are horizontal as **Figure 5.4** shows. While perfectly elastic supply curves are for the most part unrealistic, goods with readily available inputs and whose production can easily expand will feature highly elastic supply curves. Examples include pizza, bread, books, and pencils. Similarly, perfectly elastic demand is an extreme example. However, luxury goods, items that take a large share of individuals' income, and goods with many substitutes are likely to have highly elastic demand curves. Examples of such goods are Caribbean cruises and sports vehicles.

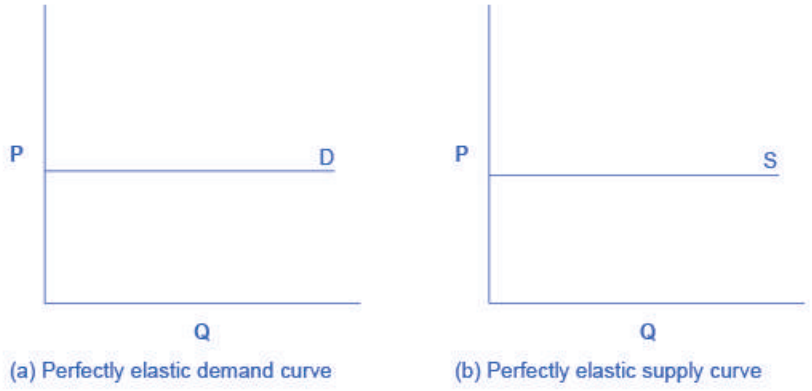


Figure 5.4 Infinite Elasticity The horizontal lines show that an infinite quantity will be demanded or supplied at a specific price. This illustrates the cases of a perfectly (or infinitely) elastic demand curve and supply curve. The quantity supplied or demanded is extremely responsive to price changes, moving from zero for prices close to P to infinite when prices reach P.

Zero elasticity or **perfect inelasticity**, as **Figure 5.5** depicts, refers to the extreme case in which a percentage change in price, no matter how large, results in zero change in quantity. While a perfectly inelastic supply is an extreme example, goods with limited supply of inputs are likely to feature highly inelastic supply curves. Examples include diamond rings or housing in prime locations such as apartments facing Central Park in New York City. Similarly,

while perfectly inelastic demand is an extreme case, necessities with no close substitutes are likely to have highly inelastic demand curves. This is the case of life-saving drugs and gasoline.

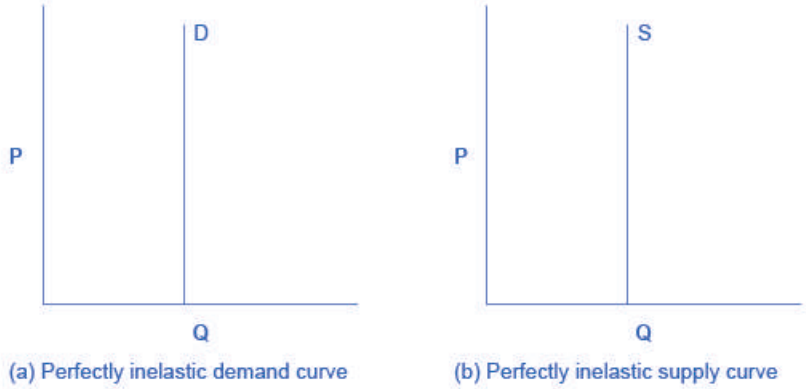


Figure 5.5 Zero Elasticity The vertical supply curve and vertical demand curve show that there will be zero percentage change in quantity (a) demanded or (b) supplied, regardless of the price.

Constant unitary elasticity, in either a supply or demand curve, occurs when a price change of one percent results in a quantity change of one percent. **Figure 5.6** shows a demand curve with constant unit elasticity. Constant unitary elasticity, in either a supply or demand curve, occurs when a price change of one percent results in a quantity change of one percent. Figure 5.6 shows a demand curve with constant unit elasticity. Using the midpoint method, you can calculate that between points A and B on the demand curve, the price changes by 28.6% and quantity demanded also changes by 28.6%. Hence, the elasticity equals 1. Between points B and C, price again changes by 28.6% as does quantity, while between points C and D the corresponding percentage changes are 22.2% for both price and quantity. In each case, then, the percentage change in price equals the percentage change in quantity, and consequently elasticity equals 1. Notice that in absolute value, the declines in price, as you step down the demand curve, are not identical. Instead, the price falls by $2.00 from A to B, by a smaller amount of $1.50 from B to C, and by a still smaller amount of $0.90 from C to D. As a result, a demand curve with constant unitary elasticity moves from a steeper slope on the left and a flatter slope on the right—and a curved shape overall. Notice that in absolute value, the declines in price, as you step down the demand curve, are not identical. Instead, the price falls by $23 from A to B, by a smaller amount of $1.50 from B to C, and by a still smaller amount of $.90 from C to D. As a result, a demand curve with constant unitary elasticity has a steeper slope on the left and a flatter slope on the right—and a curved shape overall.

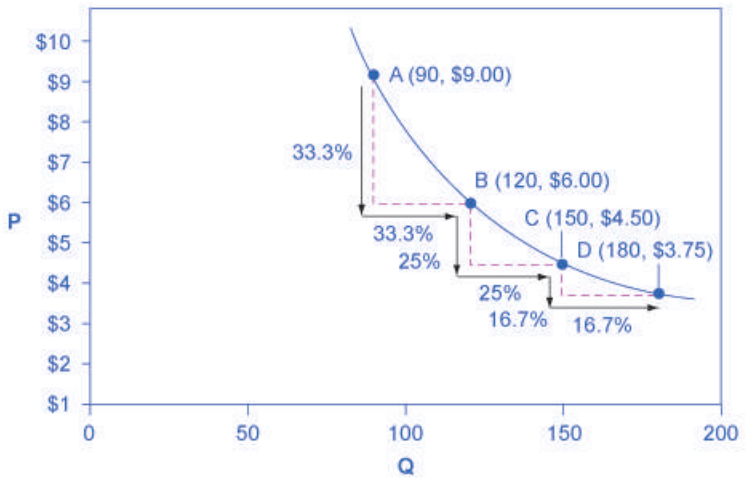


Figure 5.6 A Constant Unitary Elasticity Demand Curve A demand curve with constant unitary elasticity will be a curved line. Notice how price and quantity demanded change by an identical percentage amount between each pair of points on the demand curve.

Unlike the demand curve with unitary elasticity, the supply curve with unitary elasticity is represented by a straight

line, and that line goes through the origin. In each pair of points on the supply curve there is an equal difference in quantity of 30. However, in percentage value, using the midpoint method, the steps are decreasing as one moves from left to right, from 28.6% to 22.2% to 18.2%, because the quantity points in each percentage calculation are getting increasingly larger, which expands the denominator in the elasticity calculation of the percentage change in quantity.

Consider the price changes moving up the supply curve in **Figure 5.7**. From points D to E to F and to G on the supply curve, each step of $1.50 is the same in absolute value. However, if we measure the price changes in percentage change terms, using the midpoint method, they are also decreasing, from 28.6% to 22.2% to 18.2%, because the original price points in each percentage calculation are getting increasingly larger in value, increasing the denominator in the calculation of the percentage change in price. Along the constant unitary elasticity supply curve, the percentage quantity increases on the horizontal axis exactly match the percentage price increases on the vertical axis—so this supply curve has a constant unitary elasticity at all points.

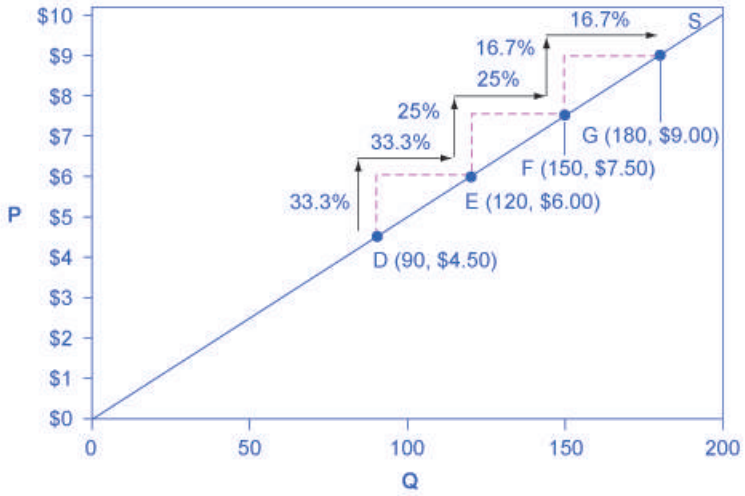


Figure 5.7 A Constant Unitary Elasticity Supply Curve A constant unitary elasticity supply curve is a straight line reaching up from the origin. Between each pair of points, the percentage increase in quantity supplied is the same as the percentage increase in price.

5.3 | Elasticity and Pricing

By the end of this section, you will be able to:

- Analyze how price elasticities impact revenue
- Evaluate how elasticity can cause shifts in demand and supply
- Predict how the long-run and short-run impacts of elasticity affect equilibrium
- Explain how the elasticity of demand and supply determine the incidence of a tax on buyers and sellers

Studying elasticities is useful for a number of reasons, pricing being most important. Let's explore how elasticity relates to revenue and pricing, both in the long and short run. First, let's look at the elasticities of some common goods and services.

Table 5.2 shows a selection of demand elasticities for different goods and services drawn from a variety of different studies by economists, listed in order of increasing elasticity.

Goods and Services	Elasticity of Price
Housing	0.12
Transatlantic air travel (economy class)	0.12
Rail transit (rush hour)	0.15
Electricity	0.20
Taxi cabs	0.22
Gasoline	0.35
Transatlantic air travel (first class)	0.40
Wine	0.55
Beef	0.59
Transatlantic air travel (business class)	0.62
Kitchen and household appliances	0.63
Cable TV (basic rural)	0.69
Chicken	0.64
Soft drinks	0.70
Beer	0.80
New vehicle	0.87
Rail transit (off-peak)	1.00
Computer	1.44
Cable TV (basic urban)	1.51
Cable TV (premium)	1.77
Restaurant meals	2.27

Table 5.2 Some Selected Elasticities of Demand

Note that demand for necessities such as housing and electricity is inelastic, while items that are not necessities such as restaurant meals are more price-sensitive. If the price of a restaurant meal increases by 10%, the quantity demanded will decrease by 22.7%. A 10% increase in the price of housing will cause only a slight decrease of 1.2% in the quantity of housing demanded.

Link It Up

Read this article (http://openstaxcollege.org/l/Movietickets) for an example of price elasticity that may have affected you.

Does Raising Price Bring in More Revenue?

Imagine that a band on tour is playing in an indoor arena with 15,000 seats. To keep this example simple, assume that the band keeps all the money from ticket sales. Assume further that the band pays the costs for its appearance, but that these costs, like travel, and setting up the stage, are the same regardless of how many people are in the audience. Finally, assume that all the tickets have the same price. (The same insights apply if ticket prices are more expensive for some seats than for others, but the calculations become more complicated.) The band knows that it faces a downward-sloping demand curve; that is, if the band raises the ticket price and, it will sell fewer seats. How should the band set the ticket price to generate the most total revenue, which in this example, because costs are fixed, will also mean the highest profits for the band? Should the band sell more tickets at a lower price or fewer tickets at a higher price?

The key concept in thinking about collecting the most revenue is the price elasticity of demand. Total revenue is price times the quantity of tickets sold. Imagine that the band starts off thinking about a certain price, which will result in the sale of a certain quantity of tickets. The three possibilities are in **Table 5.3**. If demand is elastic at that price level, then the band should cut the price, because the percentage drop in price will result in an even larger percentage increase in the quantity sold—thus raising total revenue. However, if demand is inelastic at that original quantity level, then the band should raise the ticket price, because a certain percentage increase in price will result in a smaller percentage decrease in the quantity sold—and total revenue will rise. If demand has a unitary elasticity at that quantity, then an equal percentage change in quantity will offset a moderate percentage change in the price—so the band will earn the same revenue whether it (moderately) increases or decreases the ticket price.

If Demand Is . . .	Then . . .	Therefore . . .
Elastic	% change in Qd > % change in P	A given % rise in P will be more than offset by a larger % fall in Q so that total revenue (P × Q) falls.
Unitary	% change in Qd = % change in P	A given % rise in P will be exactly offset by an equal % fall in Q so that total revenue (P × Q) is unchanged.
Inelastic	% change in Qd < % change in P	A given % rise in P will cause a smaller % fall in Q so that total revenue (P × Q) rises.

Table 5.3 Will the Band Earn More Revenue by Changing Ticket Prices?

What if the band keeps cutting price, because demand is elastic, until it reaches a level where it sells all 15,000 seats in the available arena? If demand remains elastic at that quantity, the band might try to move to a bigger arena, so that it could slash ticket prices further and see a larger percentage increase in the quantity of tickets sold. However, if the 15,000-seat arena is all that is available or if a larger arena would add substantially to costs, then this option may not work.

Conversely, a few bands are so famous, or have such fanatical followings, that demand for tickets may be inelastic right up to the point where the arena is full. These bands can, if they wish, keep raising the ticket price. Ironically,

some of the most popular bands could make more revenue by setting prices so high that the arena is not full—but those who buy the tickets would have to pay very high prices. However, bands sometimes choose to sell tickets for less than the absolute maximum they might be able to charge, often in the hope that fans will feel happier and spend more on recordings, T-shirts, and other paraphernalia.

Can Businesses Pass Costs on to Consumers?

Most businesses face a day-to-day struggle to figure out ways to produce at a lower cost, as one pathway to their goal of earning higher profits. However, in some cases, the price of a key input over which the firm has no control may rise. For example, many chemical companies use petroleum as a key input, but they have no control over the world market price for crude oil. Coffee shops use coffee as a key input, but they have no control over the world market price of coffee. If the cost of a key input rises, can the firm pass those higher costs along to consumers in the form of higher prices? Conversely, if new and less expensive ways of producing are invented, can the firm keep the benefits in the form of higher profits, or will the market pressure them to pass the gains along to consumers in the form of lower prices? The price elasticity of demand plays a key role in answering these questions.

Imagine that as a consumer of legal pharmaceutical products, you read a newspaper story that a technological breakthrough in the production of aspirin has occurred, so that every aspirin factory can now produce aspirin more cheaply. What does this discovery mean to you? **Figure 5.8** illustrates two possibilities. In **Figure 5.8** (a), the demand curve is highly inelastic. In this case, a technological breakthrough that shifts supply to the right, from S_0 to S_1, so that the equilibrium shifts from E_0 to E_1, creates a substantially lower price for the product with relatively little impact on the quantity sold. In **Figure 5.8** (b), the demand curve is highly elastic. In this case, the technological breakthrough leads to a much greater quantity sold in the market at very close to the original price. Consumers benefit more, in general, when the demand curve is more inelastic because the shift in the supply results in a much lower price for consumers.

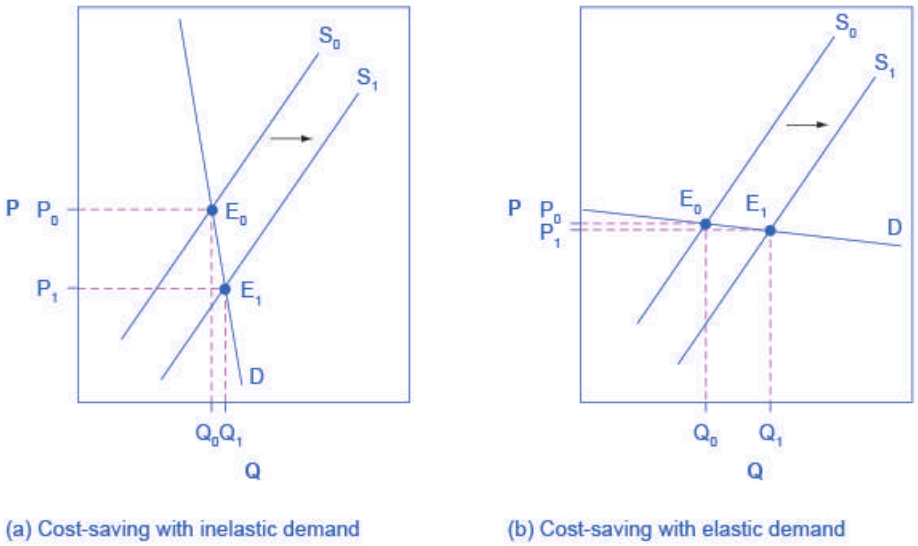


Figure 5.8 Passing along Cost Savings to Consumers Cost-saving gains cause supply to shift out to the right from S_0 to S_1; that is, at any given price, firms will be willing to supply a greater quantity. If demand is inelastic, as in (a), the result of this cost-saving technological improvement will be substantially lower prices. If demand is elastic, as in (b), the result will be only slightly lower prices. Consumers benefit in either case, from a greater quantity at a lower price, but the benefit is greater when demand is inelastic, as in (a).

Aspirin producers may find themselves in a nasty bind here. The situation in **Figure 5.8**, with extremely inelastic demand, means that a new invention may cause the price to drop dramatically while quantity changes little. As a result, the new production technology can lead to a drop in the revenue that firms earn from aspirin sales. However, if strong competition exists between aspirin producer, each producer may have little choice but to search for and implement any breakthrough that allows it to reduce production costs. After all, if one firm decides not to implement such a cost-saving technology, other firms that do can drive them out of business.

Since demand for food is generally inelastic, farmers may often face the situation in **Figure 5.8** (a). That is, a surge in

production leads to a severe drop in price that can actually decrease the total revenue that farmers receive. Conversely, poor weather or other conditions that cause a terrible year for farm production can sharply raise prices so that the total revenue that the farmer receives increases. The Clear It Up box discusses how these issues relate to coffee.

Clear It Up

How do coffee prices fluctuate?

Coffee is an international crop. The top five coffee-exporting nations are Brazil, Vietnam, Colombia, Indonesia, and Ethiopia. In these nations and others, 20 million families depend on selling coffee beans as their main source of income. These families are exposed to enormous risk, because the world price of coffee bounces up and down. For example, in 1993, the world price of coffee was about 50 cents per pound. In 1995 it was four times as high, at $2 per pound. By 1997 it had fallen by half to $1.00 per pound. In 1998 it leaped back up to $2 per pound. By 2001 it had fallen back to 46 cents a pound. By early 2011 it rose to about $2.31 per pound. By the end of 2012, the price had fallen back to about $1.31 per pound.

The reason for these price fluctuations lies in a combination of inelastic demand and shifts in supply. The elasticity of coffee demand is only about 0.3; that is, a 10% rise in the price of coffee leads to a decline of about 3% in the quantity of coffee consumed. When a major frost hit the Brazilian coffee crop in 1994, coffee supply shifted to the left with an inelastic demand curve, leading to much higher prices. Conversely, when Vietnam entered the world coffee market as a major producer in the late 1990s, the supply curve shifted out to the right. With a highly inelastic demand curve, coffee prices fell dramatically. **Figure 5.8** (a) illustrates this situation.

Elasticity also reveals whether firms can pass higher costs that they incur on to consumers. Addictive substances, for which demand is inelastic, are products for which producers can pass higher costs on to consumers. For example, the demand for cigarettes is relatively inelastic among regular smokers who are somewhat addicted. Economic research suggests that increasing cigarette prices by 10% leads to about a 3% reduction in the quantity of cigarettes that adults smoke, so the elasticity of demand for cigarettes is 0.3. If society increases taxes on companies that produce cigarettes, the result will be, as in **Figure 5.9** (a), that the supply curve shifts from S_0 to S_1. However, as the equilibrium moves from E_0 to E_1, governments mainly pass along these taxes to consumers in the form of higher prices. These higher taxes on cigarettes will raise tax revenue for the government, but they will not much affect the quantity of smoking.

If the goal is to reduce the quantity of cigarettes demanded, we must achieve it by shifting this inelastic demand back to the left, perhaps with public programs to discourage cigarette use or to help people to quit. For example, anti-smoking advertising campaigns have shown some ability to reduce smoking. However, if cigarette demand were more elastic, as in **Figure 5.9** (b), then an increase in taxes that shifts supply from S_0 to S_1 and equilibrium from E_0 to E_1 would reduce the quantity of cigarettes smoked substantially. Youth smoking seems to be more elastic than adult smoking—that is, the quantity of youth smoking will fall by a greater percentage than the quantity of adult smoking in response to a given percentage increase in price.

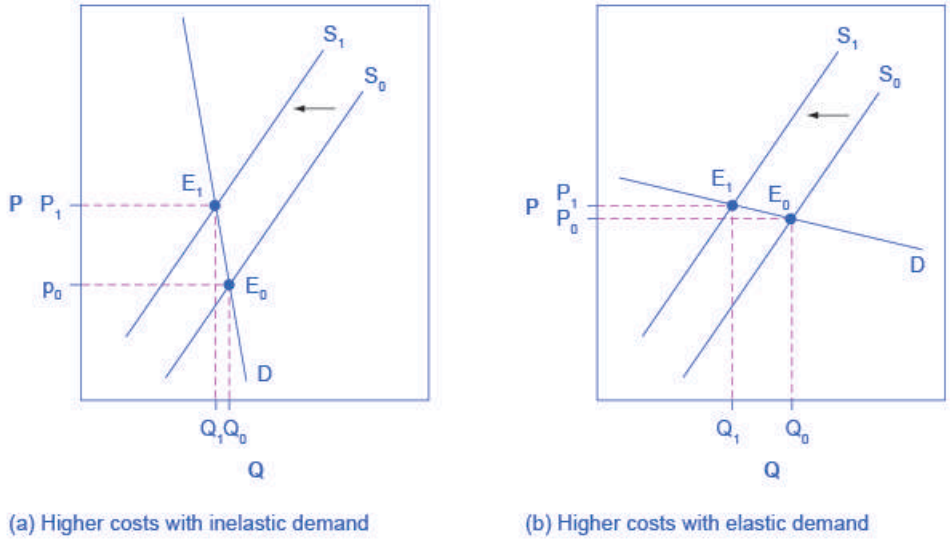


Figure 5.9 Passing along Higher Costs to Consumers Higher costs, like a higher tax on cigarette companies for the example we gave in the text, lead supply to shift to the left. This shift is identical in (a) and (b). However, in (a), where demand is inelastic, companies largely can pass the cost increase along to consumers in the form of higher prices, without much of a decline in equilibrium quantity. In (b), demand is elastic, so the shift in supply results primarily in a lower equilibrium quantity. Consumers suffer in either case, but in (a), they suffer from paying a higher price for the same quantity, while in (b), they suffer from buying a lower quantity (and presumably needing to shift their consumption elsewhere).

Elasticity and Tax Incidence

The example of cigarette taxes demonstrated that because demand is inelastic, taxes are not effective at reducing the equilibrium quantity of smoking, and they mainly pass along to consumers in the form of higher prices. The analysis, or manner, of how a tax burden is divided between consumers and producers is called **tax incidence**. Typically, the tax incidence, or burden, falls both on the consumers and producers of the taxed good. However, if one wants to predict which group will bear most of the burden, all one needs to do is examine the elasticity of demand and supply. In the tobacco example, the tax burden falls on the most inelastic side of the market.

If demand is more inelastic than supply, consumers bear most of the tax burden, and if supply is more inelastic than demand, sellers bear most of the tax burden.

The intuition for this is simple. When the demand is inelastic, consumers are not very responsive to price changes, and the quantity demanded reduces only modestly when the tax is introduced. In the case of smoking, the demand is inelastic because consumers are addicted to the product. The government can then pass the tax burden along to consumers in the form of higher prices, without much of a decline in the equilibrium quantity.

Similarly, when a government introduces a tax in a market with an inelastic supply, such as, for example, beachfront hotels, and sellers have no alternative than to accept lower prices for their business, taxes do not greatly affect the equilibrium quantity. The tax burden now passes on to the sellers. If the supply was elastic and sellers had the possibility of reorganizing their businesses to avoid supplying the taxed good, the tax burden on the sellers would be much smaller. The tax would result in a much lower quantity sold instead of lower prices received. **Figure 5.10** illustrates this relationship between the tax incidence and elasticity of demand and supply.

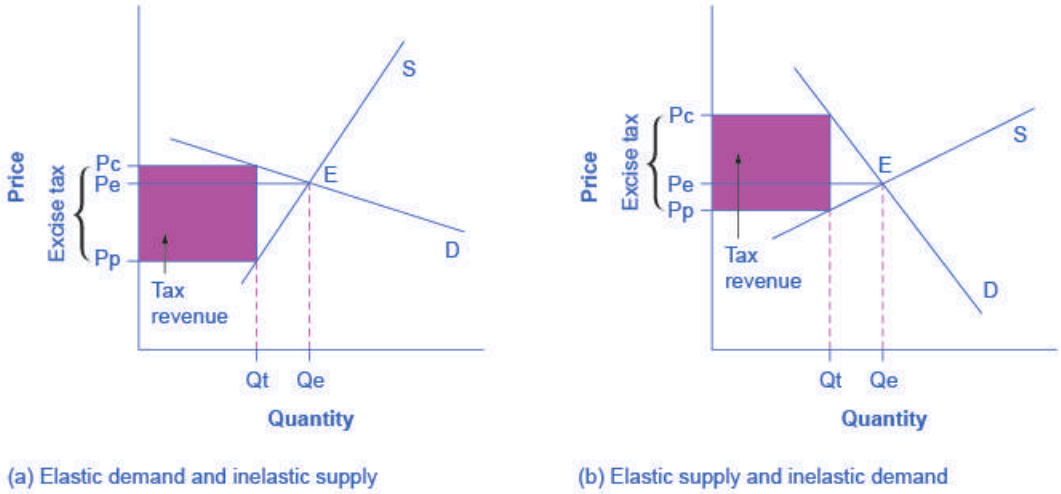


Figure 5.10 Elasticity and Tax Incidence An excise tax introduces a wedge between the price paid by consumers (Pc) and the price received by producers (Pp). The vertical distance between Pc and Pp is the amount of the tax per unit. Pe is the equilibrium price prior to introduction of the tax. (a) When the demand is more elastic than supply, the tax incidence on consumers Pc – Pe is lower than the tax incidence on producers Pe – Pp. (b) When the supply is more elastic than demand, the tax incidence on consumers Pc – Pe is larger than the tax incidence on producers Pe – Pp. The more elastic the demand and supply curves, the lower the tax revenue.

In **Figure 5.10** (a), the supply is inelastic and the demand is elastic, such as in the example of beachfront hotels. While consumers may have other vacation choices, sellers can't easily move their businesses. By introducing a tax, the government essentially creates a wedge between the price paid by consumers Pc and the price received by producers Pp. In other words, of the total price paid by consumers, part is retained by the sellers and part is paid to the government in the form of a tax. The distance between Pc and Pp is the tax rate. The new market price is Pc, but sellers receive only Pp per unit sold, as they pay Pc-Pp to the government. Since we can view a tax as raising the costs of production, this could also be represented by a leftward shift of the supply curve, where the new supply curve would intercept the demand at the new quantity Qt. For simplicity, **Figure 5.10** omits the shift in the supply curve.

The tax revenue is given by the shaded area, which we obtain by multiplying the tax per unit by the total quantity sold Qt. The tax incidence on the consumers is given by the difference between the price paid Pc and the initial equilibrium price Pe. The tax incidence on the sellers is given by the difference between the initial equilibrium price Pe and the price they receive after the tax is introduced Pp. In **Figure 5.10** (a), the tax burden falls disproportionately on the sellers, and a larger proportion of the tax revenue (the shaded area) is due to the resulting lower price received by the sellers than by the resulting higher prices paid by the buyers. **Figure 5.10** (b) describes the example of the tobacco excise tax where the supply is more elastic than demand. The tax incidence now falls disproportionately on consumers, as shown by the large difference between the price they pay, Pc, and the initial equilibrium price, Pe. Sellers receive a lower price than before the tax, but this difference is much smaller than the change in consumers' price. From this analysis one can also predict whether a tax is likely to create a large revenue or not. The more elastic the demand curve, the more likely that consumers will reduce quantity instead of paying higher prices. The more elastic the supply curve, the more likely that sellers will reduce the quantity sold, instead of taking lower prices. In a market where both the demand and supply are very elastic, the imposition of an excise tax generates low revenue.

Some believe that excise taxes hurt mainly the specific industries they target. For example, the medical device excise tax, in effect since 2013, has been controversial for it can delay industry profitability and therefore hamper start-ups and medical innovation. However, whether the tax burden falls mostly on the medical device industry or on the patients depends simply on the elasticity of demand and supply.

Long-Run vs. Short-Run Impact

Elasticities are often lower in the short run than in the long run. On the demand side of the market, it can sometimes be difficult to change Qd in the short run, but easier in the long run. Consumption of energy is a clear example. In the short run, it is not easy for a person to make substantial changes in energy consumption. Maybe you can carpool to work sometimes or adjust your home thermostat by a few degrees if the cost of energy rises, but that is about all.

However, in the long run you can purchase a car that gets more miles to the gallon, choose a job that is closer to where you live, buy more energy-efficient home appliances, or install more insulation in your home. As a result, the elasticity of demand for energy is somewhat inelastic in the short run, but much more elastic in the long run.

Figure 5.11 is an example, based roughly on historical experience, for the responsiveness of Qd to price changes. In 1973, the price of crude oil was $12 per barrel and total consumption in the U.S. economy was 17 million barrels per day. That year, the nations who were members of the Organization of Petroleum Exporting Countries (OPEC) cut off oil exports to the United States for six months because the Arab members of OPEC disagreed with the U.S. support for Israel. OPEC did not bring exports back to their earlier levels until 1975—a policy that we can interpret as a shift of the supply curve to the left in the U.S. petroleum market. **Figure 5.11** (a) and **Figure 5.11** (b) show the same original equilibrium point and the same identical shift of a supply curve to the left from S_0 to S_1.

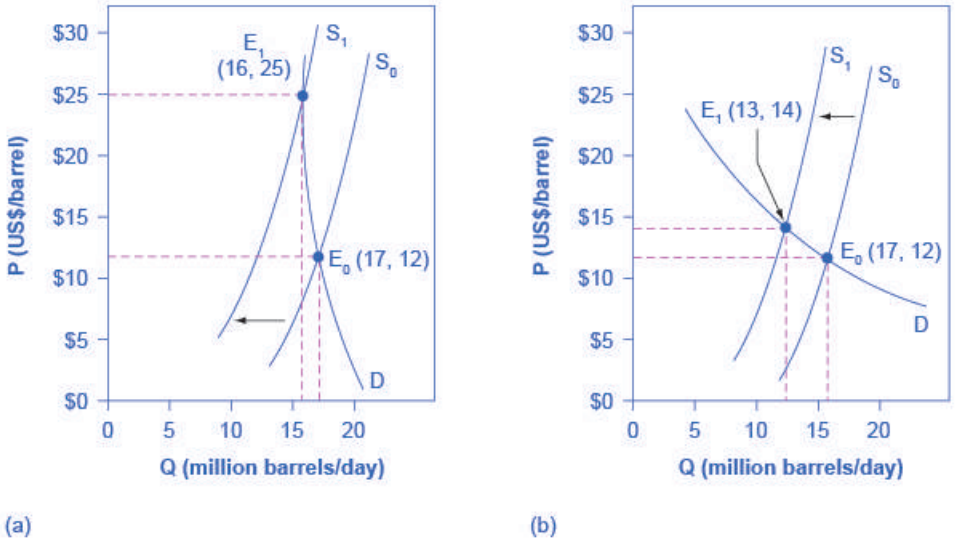


Figure 5.11 How a Shift in Supply Can Affect Price or Quantity The intersection (E_0) between demand curve D and supply curve S_0 is the same in both (a) and (b). The shift of supply to the left from S_0 to S_1 is identical in both (a) and (b). The new equilibrium (E_1) has a higher price and a lower quantity than the original equilibrium (E_0) in both (a) and (b). However, the shape of the demand curve D is different in (a) and (b), being more elastic in (b) than in (a). As a result, the shift in supply can result either in a new equilibrium with a much higher price and an only slightly smaller quantity, as in (a), with more inelastic demand, or in a new equilibrium with only a small increase in price and a relatively larger reduction in quantity, as in (b), with more elastic demand.

Figure 5.11 (a) shows inelastic demand for oil in the short run similar to that which existed for the United States in 1973. In **Figure 5.11** (a), the new equilibrium (E_1) occurs at a price of $25 per barrel, roughly double the price before the OPEC shock, and an equilibrium quantity of 16 million barrels per day. **Figure 5.11** (b) shows what the outcome would have been if the U.S. demand for oil had been more elastic, a result more likely over the long term. This alternative equilibrium (E_1) would have resulted in a smaller price increase to $14 per barrel and larger reduction in equilibrium quantity to 13 million barrels per day. In 1983, for example, U.S. petroleum consumption was 15.3 million barrels a day, which was lower than in 1973 or 1975. U.S. petroleum consumption was down even though the U.S. economy was about one-fourth larger in 1983 than it had been in 1973. The primary reason for the lower quantity was that higher energy prices spurred conservation efforts, and after a decade of home insulation, more fuel-efficient cars, more efficient appliances and machinery, and other fuel-conserving choices, the demand curve for energy had become more elastic.

On the supply side of markets, producers of goods and services typically find it easier to expand production in the long term of several years rather than in the short run of a few months. After all, in the short run it can be costly or difficult to build a new factory, hire many new workers, or open new stores. However, over a few years, all of these are possible.

In most markets for goods and services, prices bounce up and down more than quantities in the short run, but quantities often move more than prices in the long run. The underlying reason for this pattern is that supply and demand are often inelastic in the short run, so that shifts in either demand or supply can cause a relatively greater

change in prices. However, since supply and demand are more elastic in the long run, the long-run movements in prices are more muted, while quantity adjusts more easily in the long run.

5.4 | Elasticity in Areas Other Than Price

By the end of this section, you will be able to:

- Calculate the income elasticity of demand and the cross-price elasticity of demand
- Calculate the elasticity in labor and financial capital markets through an understanding of the elasticity of labor supply and the elasticity of savings
- Apply concepts of price elasticity to real-world situations

The basic idea of elasticity—how a percentage change in one variable causes a percentage change in another variable—does not just apply to the responsiveness quantity supplied and quantity demanded to changes in the price of a product. Recall that quantity demanded (Qd) depends on income, tastes and preferences, the prices of related goods, and so on, as well as price. Similarly, quantity supplied (Qs) depends on factors such as the cost of production, as well as price. We can measure elasticity for any determinant of quantity supplied and quantity demanded, not just the price.

Income Elasticity of Demand

The income elasticity of demand is the percentage change in quantity demanded divided by the percentage change in income.

$$\text{Income elasticity of demand} = \frac{\%\text{ change in quantity demanded}}{\%\text{ change in income}}$$

For most products, most of the time, the income elasticity of demand is positive: that is, a rise in income will cause an increase in the quantity demanded. This pattern is common enough that we refer to these goods as normal goods. However, for a few goods, an increase in income means that one might purchase less of the good. For example, those with a higher income might buy fewer hamburgers, because they are buying more steak instead, or those with a higher income might buy less cheap wine and more imported beer. When the income elasticity of demand is negative, we call the good an inferior good.

We introduced the concepts of normal and inferior goods in **Demand and Supply**. A higher level of income causes a demand curve to shift to the right for a normal good, which means that the income elasticity of demand is positive. How far the demand shifts depends on the income elasticity of demand. A higher income elasticity means a larger shift. However, for an inferior good, that is, when the income elasticity of demand is negative, a higher level of income would cause the demand curve for that good to shift to the left. Again, how much it shifts depends on how large the (negative) income elasticity is.

Cross-Price Elasticity of Demand

A change in the price of one good can shift the quantity demanded for another good. If the two goods are complements, like bread and peanut butter, then a drop in the price of one good will lead to an increase in the quantity demanded of the other good. However, if the two goods are substitutes, like plane tickets and train tickets, then a drop in the price of one good will cause people to substitute toward that good, and to reduce consumption of the other good. Cheaper plane tickets lead to fewer train tickets, and vice versa.

The **cross-price elasticity of demand** puts some meat on the bones of these ideas. The term “cross-price” refers to the idea that the price of one good is affecting the quantity demanded of a different good. Specifically, the cross-price elasticity of demand is the percentage change in the quantity of good A that is demanded as a result of a percentage change in the price of good B.

$$\text{Cross-price elasticity of demand} = \frac{\%\text{ change in Qd of good A}}{\%\text{ change in price of good B}}$$

Substitute goods have positive cross-price elasticities of demand: if good A is a substitute for good B, like coffee and tea, then a higher price for B will mean a greater quantity consumed of A. Complement goods have negative cross-price elasticities: if good A is a complement for good B, like coffee and sugar, then a higher price for B will mean a

lower quantity consumed of A.

Elasticity in Labor and Financial Capital Markets

The concept of elasticity applies to any market, not just markets for goods and services. In the labor market, for example, the **wage elasticity of labor supply**—that is, the percentage change in hours worked divided by the percentage change in wages—will reflect the shape of the labor supply curve. Specifically:

$$\text{Elasticity of labor supply} = \frac{\%\text{ change in quantity of labor supplied}}{\%\text{ change in wage}}$$

The wage elasticity of labor supply for teenage workers is generally fairly elastic: that is, a certain percentage change in wages will lead to a larger percentage change in the quantity of hours worked. Conversely, the wage elasticity of labor supply for adult workers in their thirties and forties is fairly inelastic. When wages move up or down by a certain percentage amount, the quantity of hours that adults in their prime earning years are willing to supply changes but by a lesser percentage amount.

In markets for financial capital, the **elasticity of savings**—that is, the percentage change in the quantity of savings divided by the percentage change in interest rates—will describe the shape of the supply curve for financial capital. That is:

$$\text{Elasticity of savings} = \frac{\%\text{ change in quantity of financial savings}}{\%\text{ change in interest rate}}$$

Sometimes laws are proposed that seek to increase the quantity of savings by offering tax breaks so that the return on savings is higher. Such a policy will have a comparatively large impact on increasing the quantity saved if the supply curve for financial capital is elastic, because then a given percentage increase in the return to savings will cause a higher percentage increase in the quantity of savings. However, if the supply curve for financial capital is highly inelastic, then a percentage increase in the return to savings will cause only a small increase in the quantity of savings. The evidence on the supply curve of financial capital is controversial but, at least in the short run, the elasticity of savings with respect to the interest rate appears fairly inelastic.

Expanding the Concept of Elasticity

The elasticity concept does not even need to relate to a typical supply or demand curve at all. For example, imagine that you are studying whether the Internal Revenue Service should spend more money on auditing tax returns. We can frame the question in terms of the elasticity of tax collections with respect to spending on tax enforcement; that is, what is the percentage change in tax collections derived from a given percentage change in spending on tax enforcement?

With all of the elasticity concepts that we have just described, some of which are in **Table 5.4**, the possibility of confusion arises. When you hear the phrases “elasticity of demand” or “elasticity of supply,” they refer to the elasticity with respect to price. Sometimes, either to be extremely clear or because economists are discussing a wide variety of elasticities, we will call the elasticity of demand or the demand elasticity the price elasticity of demand or the “elasticity of demand with respect to price.” Similarly, economists sometimes use the term elasticity of supply or the supply elasticity, to avoid any possibility of confusion, the price elasticity of supply or “the elasticity of supply with respect to price.” However, in whatever context, the idea of elasticity always refers to percentage change in one variable, almost always a price or money variable, and how it causes a percentage change in another variable, typically a quantity variable of some kind.

Income elasticity of demand	$= \frac{\%\text{ change in Qd}}{\%\text{ change in income}}$
Cross-price elasticity of demand	$= \frac{\%\text{ change in Qd of good A}}{\%\text{ change in price of good B}}$

Table 5.4 Formulas for Calculating Elasticity

Wage elasticity of labor supply	$= \frac{\%\text{ change in quantity of labor supplied}}{\%\text{ change in wage}}$
Wage elasticity of labor demand	$= \frac{\%\text{ change in quantity of labor demanded}}{\%\text{ change in wage}}$
Interest rate elasticity of savings	$= \frac{\%\text{ change in quantity of savings}}{\%\text{ change in interest rate}}$
Interest rate elasticity of borrowing	$= \frac{\%\text{ change in quantity of borrowing}}{\%\text{ change in interest rate}}$

Table 5.4 Formulas for Calculating Elasticity

Bring it Home

That Will Be How Much?

How did the 60% price increase in 2011 end up for Netflix? It has been a very bumpy ride.

Before the price increase, there were about 24.6 million U.S. subscribers. After the price increase, 810,000 infuriated U.S. consumers canceled their Netflix subscriptions, dropping the total number of subscribers to 23.79 million. Fast forward to June 2013, when there were 36 million streaming Netflix subscribers in the United States. This was an increase of 11.4 million subscribers since the price increase—an average per quarter growth of about 1.6 million. This growth is less than the 2 million per quarter increases Netflix experienced in the fourth quarter of 2010 and the first quarter of 2011.

During the first year after the price increase, the firm's stock price (a measure of future expectations for the firm) fell from about $33.60 per share per share to just under $7.80. By the end of 2016, however, the stock price was at $123 per share. Today, Netflix has more than 86 million subscribers million subscribers in fifty countries.

What happened? Obviously, Netflix company officials understood the law of demand. Company officials reported, when announcing the price increase, this could result in the loss of about 600,000 existing subscribers. Using the elasticity of demand formula, it is easy to see company officials expected an inelastic response:

$$\begin{aligned} &= \frac{-600{,}000/[(24\text{ million} + 24.6\text{ million})/2]}{\$6/[(\$10 + \$16)/2]} \\ &= \frac{-600{,}000/24.3\text{ million}}{\$6/\$13} \\ &= \frac{-0.025}{0.46} \\ &= -0.05 \end{aligned}$$

In addition, Netflix officials had anticipated the price increase would have little impact on attracting new customers. Netflix anticipated adding up to 1.29 million new subscribers in the third quarter of 2011. It is true this was slower growth than the firm had experienced—about 2 million per quarter.

Why was the estimate of customers leaving so far off? In the more than two decades since Netflix had been founded, there was an increase in the number of close, but not perfect, substitutes. Consumers now had choices ranging from Vudu, Amazon Prime, Hulu, and Redbox, to retail stores. Jaime Weinman reported in *Maclean's* that Redbox kiosks are "a five-minute drive for less from 68 percent of Americans, and it seems that many people still find a five-minute drive more convenient than loading up a movie online." It seems that in 2012, many consumers still preferred a physical DVD disk over streaming video.

What missteps did the Netflix management make? In addition to misjudging the elasticity of demand, by failing

to account for close substitutes, it seems they may have also misjudged customers' preferences and tastes. Yet, as the population increases, the preference for streaming video may overtake physical DVD disks. Netflix, the source of numerous late night talk show laughs and jabs in 2011, may yet have the last laugh.

KEY TERMS

constant unitary elasticity when a given percent price change in price leads to an equal percentage change in quantity demanded or supplied

cross-price elasticity of demand the percentage change in the quantity of good A that is demanded as a result of a percentage change in good B

elastic demand when the elasticity of demand is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elastic supply when the elasticity of either supply is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elasticity an economics concept that measures responsiveness of one variable to changes in another variable

elasticity of savings the percentage change in the quantity of savings divided by the percentage change in interest rates

inelastic demand when the elasticity of demand is less than one, indicating that a 1 percent increase in price paid by the consumer leads to less than a 1 percent change in purchases (and vice versa); this indicates a low responsiveness by consumers to price changes

inelastic supply when the elasticity of supply is less than one, indicating that a 1 percent increase in price paid to the firm will result in a less than 1 percent increase in production by the firm; this indicates a low responsiveness of the firm to price increases (and vice versa if prices drop)

infinite elasticity the extremely elastic situation of demand or supply where quantity changes by an infinite amount in response to any change in price; horizontal in appearance

perfect elasticity see infinite elasticity

perfect inelasticity see zero elasticity

price elasticity the relationship between the percent change in price resulting in a corresponding percentage change in the quantity demanded or supplied

price elasticity of demand percentage change in the quantity *demanded* of a good or service divided the percentage change in price

price elasticity of supply percentage change in the quantity *supplied* divided by the percentage change in price

tax incidence manner in which the tax burden is divided between buyers and sellers

unitary elasticity when the calculated elasticity is equal to one indicating that a change in the price of the good or service results in a proportional change in the quantity demanded or supplied

wage elasticity of labor supply the percentage change in hours worked divided by the percentage change in wages

zero inelasticity the highly inelastic case of demand or supply in which a percentage change in price, no matter how large, results in zero change in the quantity; vertical in appearance

KEY CONCEPTS AND SUMMARY

5.1 Price Elasticity of Demand and Price Elasticity of Supply

Price elasticity measures the responsiveness of the quantity demanded or supplied of a good to a change in its price. We compute it as the percentage change in quantity demanded (or supplied) divided by the percentage change in price. We can describe elasticity as elastic (or very responsive), unit elastic, or inelastic (not very responsive). Elastic demand or supply curves indicate that quantity demanded or supplied respond to price changes in a greater than proportional manner. An inelastic demand or supply curve is one where a given percentage change in price will cause a smaller percentage change in quantity demanded or supplied. A unitary elasticity means that a given percentage change in price leads to an equal percentage change in quantity demanded or supplied.

5.2 Polar Cases of Elasticity and Constant Elasticity

Infinite or perfect elasticity refers to the extreme case where either the quantity demanded or supplied changes by an infinite amount in response to any change in price at all. Zero elasticity refers to the extreme case in which a percentage change in price, no matter how large, results in zero change in quantity. Constant unitary elasticity in either a supply or demand curve refers to a situation where a price change of one percent results in a quantity change of one percent.

5.3 Elasticity and Pricing

In the market for goods and services, quantity supplied and quantity demanded are often relatively slow to react to changes in price in the short run, but react more substantially in the long run. As a result, demand and supply often (but not always) tend to be relatively inelastic in the short run and relatively elastic in the long run. A tax incidence depends on the relative price elasticity of supply and demand. When supply is more elastic than demand, buyers bear most of the tax burden, and when demand is more elastic than supply, producers bear most of the cost of the tax. Tax revenue is larger the more inelastic the demand and supply are.

5.4 Elasticity in Areas Other Than Price

Elasticity is a general term, that reflects responsiveness. It refers to the change of one variable divided by the percentage change of a related variable that we can apply to many economic connections. For instance, the income elasticity of demand is the percentage change in quantity demanded divided by the percentage change in income. The cross-price elasticity of demand is the percentage change in the quantity demanded of a good divided by the percentage change in the price of another good. Elasticity applies in labor markets and financial capital markets just as it does in markets for goods and services. The wage elasticity of labor supply is the percentage change in the quantity of hours supplied divided by the percentage change in the wage. The elasticity of savings with respect to interest rates is the percentage change in the quantity of savings divided by the percentage change in interest rates.

SELF-CHECK QUESTIONS

1. From the data in **Table 5.5** about demand for smart phones, calculate the price elasticity of demand from: point B to point C, point D to point E, and point G to point H. Classify the elasticity at each point as elastic, inelastic, or unit elastic.

Points	P	Q
A	60	3,000
B	70	2,800
C	80	2,600
D	90	2,400
E	100	2,200
F	110	2,000
G	120	1,800
H	130	1,600

Table 5.5

2. From the data in **Table 5.6** about supply of alarm clocks, calculate the price elasticity of supply from: point J to point K, point L to point M, and point N to point P. Classify the elasticity at each point as elastic, inelastic, or unit elastic.

Point	Price	Quantity Supplied
J	$8	50
K	$9	70
L	$10	80
M	$11	88
N	$12	95
P	$13	100

Table 5.6

3. Why is the demand curve with constant unitary elasticity concave?

4. Why is the supply curve with constant unitary elasticity a straight line?

5. The federal government decides to require that automobile manufacturers install new anti-pollution equipment that costs $2,000 per car. Under what conditions can carmakers pass almost all of this cost along to car buyers? Under what conditions can carmakers pass very little of this cost along to car buyers?

6. Suppose you are in charge of sales at a pharmaceutical company, and your firm has a new drug that causes bald men to grow hair. Assume that the company wants to earn as much revenue as possible from this drug. If the elasticity of demand for your company's product at the current price is 1.4, would you advise the company to raise the price, lower the price, or to keep the price the same? What if the elasticity were 0.6? What if it were 1? Explain your answer.

7. What would the gasoline price elasticity of supply mean to UPS or FedEx?

8. The average annual income rises from $25,000 to $38,000, and the quantity of bread consumed in a year by the average person falls from 30 loaves to 22 loaves. What is the income elasticity of bread consumption? Is bread a normal or an inferior good?

9. Suppose the cross-price elasticity of apples with respect to the price of oranges is 0.4, and the price of oranges falls by 3%. What will happen to the demand for apples?

REVIEW QUESTIONS

10. What is the formula for calculating elasticity?

11. What is the price elasticity of demand? Can you explain it in your own words?

12. What is the price elasticity of supply? Can you explain it in your own words?

13. Describe the general appearance of a demand or a supply curve with zero elasticity.

14. Describe the general appearance of a demand or a supply curve with infinite elasticity.

15. If demand is elastic, will shifts in supply have a larger effect on equilibrium quantity or on price?

16. If demand is inelastic, will shifts in supply have a larger effect on equilibrium price or on quantity?

17. If supply is elastic, will shifts in demand have a larger effect on equilibrium quantity or on price?

18. If supply is inelastic, will shifts in demand have a larger effect on equilibrium price or on quantity?

19. Would you usually expect elasticity of demand or supply to be higher in the short run or in the long run? Why?

20. Under which circumstances does the tax burden fall entirely on consumers?

21. What is the formula for the income elasticity of demand?

22. What is the formula for the cross-price elasticity of demand?

23. What is the formula for the wage elasticity of labor supply?

24. What is the formula for elasticity of savings with respect to interest rates?

CRITICAL THINKING QUESTIONS

25. Transatlantic air travel in business class has an estimated elasticity of demand of 0.40 less than transatlantic air travel in economy class, with an estimated price elasticity of 0.62. Why do you think this is the case?

26. What is the relationship between price elasticity and position on the demand curve? For example, as you move up the demand curve to higher prices and lower quantities, what happens to the measured elasticity? How would you explain that?

27. Can you think of an industry (or product) with near infinite elasticity of supply in the short term? That is, what is an industry that could increase Qs almost without limit in response to an increase in the price?

28. Would you expect supply to play a more significant role in determining the price of a basic necessity like food or a luxury like perfume? Explain. *Hint*: Think about how the price elasticity of demand will differ between necessities and luxuries.

29. A city has built a bridge over a river and it decides to charge a toll to everyone who crosses. For one year, the city charges a variety of different tolls and records information on how many drivers cross the bridge. The city thus gathers information about elasticity of demand. If the city wishes to raise as much revenue as possible from the tolls, where will the city decide to charge a toll: in the inelastic portion of the demand curve, the elastic portion of the demand curve, or the unit elastic portion? Explain.

30. In a market where the supply curve is perfectly inelastic, how does an excise tax affect the price paid by consumers and the quantity bought and sold?

31. Economists define normal goods as having a positive income elasticity. We can divide normal goods into two types: Those whose income elasticity is less than one and those whose income elasticity is greater than one. Think about products that would fall into each category. Can you come up with a name for each category?

32. Suppose you could buy shoes one at a time, rather than in pairs. What do you predict the cross-price elasticity for left shoes and right shoes would be?

PROBLEMS

33. The equation for a demand curve is $P = 48 – 3Q$. What is the elasticity in moving from a quantity of 5 to a quantity of 6?

34. The equation for a demand curve is $P = 2/Q$. What is the elasticity of demand as price falls from 5 to 4? What is the elasticity of demand as the price falls from 9 to 8? Would you expect these answers to be the same?

35. The equation for a supply curve is $4P = Q$. What is the elasticity of supply as price rises from 3 to 4? What is the elasticity of supply as the price rises from 7 to 8? Would you expect these answers to be the same?

36. The equation for a supply curve is $P = 3Q – 8$. What is the elasticity in moving from a price of 4 to a price of 7?

37. The supply of paintings by Leonardo Da Vinci, who painted the *Mona Lisa* and *The Last Supper* and died in 1519, is highly inelastic. Sketch a supply and demand diagram, paying attention to the appropriate elasticities, to illustrate that demand for these paintings will determine the price.

38. Say that a certain stadium for professional football has 70,000 seats. What is the shape of the supply curve for tickets to football games at that stadium? Explain.

39. When someone's kidneys fail, the person needs to have medical treatment with a dialysis machine (unless or until they receive a kidney transplant) or they will die. Sketch a supply and demand diagram, paying attention to the appropriate elasticities, to illustrate that the supply of such dialysis machines will primarily determine the price.

40. Assume that the supply of low-skilled workers is fairly elastic, but the employers' demand for such workers is fairly inelastic. If the policy goal is to expand employment for low-skilled workers, is it better to focus on policy tools to shift the supply of unskilled labor or on tools to shift the demand for unskilled labor? What if the policy goal is to raise wages for this group? Explain your answers with supply and demand diagrams.

6 | Consumer Choices

Figure 6.1 Investment Choices We generally view higher education as a good investment, if one can afford it, regardless of the state of the economy. (Credit: modification of work by Jason Bache/Flickr Creative Commons)

Bring it Home

"Eeny, Meeny, Miney, Moe"—Making Choices

The 2008–2009 Great Recession touched families around the globe. In too many countries, workers found themselves out of a job. In developed countries, unemployment compensation provided a safety net, but families still saw a marked decrease in disposable income and had to make tough spending decisions. Of course, non-essential, discretionary spending was the first to go.

Even so, there was one particular category that saw a universal increase in spending world-wide during that time—an 18% uptick in the United States, specifically. You might guess that consumers began eating more meals at home, increasing grocery store spending; however, the Bureau of Labor Statistics' Consumer Expenditure Survey, which tracks U.S. food spending over time, showed "real total food spending by U.S. households declined five percent between 2006 and 2009." So, it was not groceries. What product would people around the world demand more of during tough economic times, and more importantly, why? (Find out at chapter's end.)

That question leads us to this chapter's topic—analyzing how consumers make choices. For most consumers, using "eeny, meeny, miney, moe" is not how they make decisions. Their decision-making processes have been educated far beyond a children's rhyme.

Introduction to Consumer Choices

In this chapter, you will learn about:

- Consumption Choices
- How Changes in Income and Prices Affect Consumption Choices
- How Consumer Choices Might Not Always be Rational

Microeconomics seeks to understand the behavior of individual economic agents such as individuals and businesses. Economists believe that we can analyze individuals' decisions, such as what goods and services to buy, as choices we make within certain budget constraints. Generally, consumers are trying to get the most for their limited budget. In economic terms they are trying to maximize total utility, or satisfaction, given their budget constraint.

Everyone has their own personal tastes and preferences. The French say: *Chacun à son goût*, or "Each to his own taste." An old Latin saying states, *De gustibus non est disputandum* or "There's no disputing about taste." If people base their decisions on their own tastes and personal preferences, however, then how can economists hope to analyze the choices consumers make?

An economic explanation for why people make different choices begins with accepting the proverbial wisdom that tastes are a matter of personal preference. However, economists also believe that the choices people make are influenced by their incomes, by the prices of goods and services they consume, and by factors like where they live. This chapter introduces the economic theory of how consumers make choices about what goods and services to buy with their limited income.

The analysis in this chapter will build on the budget constraint that we introduced in the **Choice in a World of Scarcity** chapter. This chapter will also illustrate how economic theory provides a tool to systematically look at the full range of possible consumption choices to predict how consumption responds to changes in prices or incomes. After reading this chapter, consult the appendix **Indifference Curves** to learn more about representing utility and choice through indifference curves.

6.1 | Consumption Choices

By the end of this section, you will be able to:

- Calculate total utility
- Propose decisions that maximize utility
- Explain marginal utility and the significance of diminishing marginal utility

Information on the consumption choices of Americans is available from the Consumer Expenditure Survey carried out by the U.S. Bureau of Labor Statistics. **Table 6.1** shows spending patterns for the average U.S. household. The first row shows income and, after taxes and personal savings are subtracted, it shows that, in 2015, the average U.S. household spent $48,109 on consumption. The table then breaks down consumption into various categories. The average U.S. household spent roughly one-third of its consumption on shelter and other housing expenses, another one-third on food and vehicle expenses, and the rest on a variety of items, as shown. These patterns will vary for specific households by differing levels of family income, by geography, and by preferences.

Average Household Income before Taxes	**$62,481**
Average Annual Expenditures	**$48.109**
Food at home	$3,264
Food away from home	$2,505
Housing	$16,557
Apparel and services	$1,700

Table 6.1 U.S. Consumption Choices in 2015 (Source: http://www.bls.gov/cex/csxann13.pdf)

Transportation	$7,677
Healthcare	$3,157
Entertainment	$2,504
Education	$1,074
Personal insurance and pensions	$5,357
All else: alcohol, tobacco, reading, personal care, cash contributions, miscellaneous	$3,356

Table 6.1 U.S. Consumption Choices in 2015 (Source: http://www.bls.gov/cex/csxann13.pdf)

Total Utility and Diminishing Marginal Utility

To understand how a household will make its choices, economists look at what consumers can afford, as shown in a **budget constraint (or budget line)**, and the **total utility** or satisfaction derived from those choices. In a budget constraint line, the quantity of one good is on the horizontal axis and the quantity of the other good on the vertical axis. The budget constraint line shows the various combinations of two goods that are affordable given consumer income. Consider José's situation, shown in **Figure 6.2**. José likes to collect T-shirts and watch movies.

In **Figure 6.2** we show the quantity of T-shirts on the horizontal axis while we show the quantity of movies on the vertical axis. If José had unlimited income or goods were free, then he could consume without limit. However, José, like all of us, faces a budget constraint. José has a total of $56 to spend. The price of T-shirts is $14 and the price of movies is $7. Notice that the vertical intercept of the budget constraint line is at eight movies and zero T-shirts ($56/$7=8). The horizontal intercept of the budget constraint is four, where José spends of all of his money on T-shirts and no movies ($56/14=4). The slope of the budget constraint line is rise/run or –8/4=–2. The specific choices along the budget constraint line show the combinations of affordable T-shirts and movies.

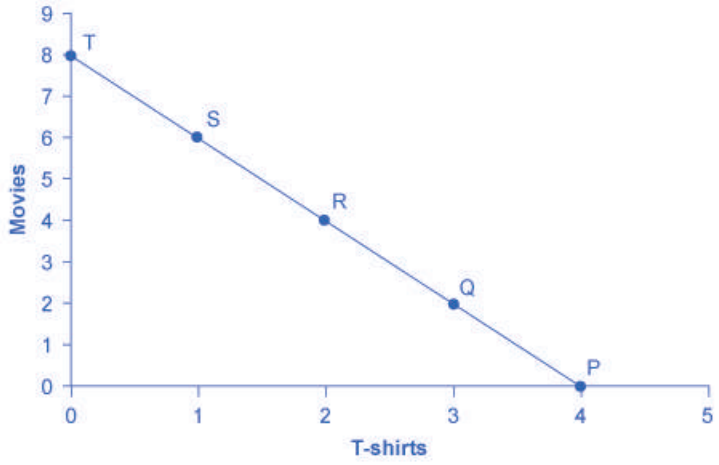


Figure 6.2 A Choice between Consumption Goods José has income of $56. Movies cost $7 and T-shirts cost $14. The points on the budget constraint line show the combinations of affordable movies and T-shirts.

José wishes to choose the combination that will provide him with the greatest utility, which is the term economists use to describe a person's level of satisfaction or happiness with his or her choices.

Let's begin with an assumption, which we will discuss in more detail later, that José can measure his own utility with something called *utils*. (It is important to note that you cannot make comparisons between the utils of individuals. If one person gets 20 utils from a cup of coffee and another gets 10 utils, this does not mean than the first person gets more enjoyment from the coffee than the other or that they enjoy the coffee twice as much. The reason why is that utils are subjective to an individual. The way one person measures utils is not the same as the way someone else does.) **Table 6.2** shows how José's utility is connected with his T-shirt or movie consumption. The first column of the table shows the quantity of T-shirts consumed. The second column shows the total utility, or total amount of

satisfaction, that José receives from consuming that number of T-shirts. The most common pattern of total utility, in this example, is that consuming additional goods leads to greater total utility, but at a decreasing rate. The third column shows **marginal utility**, which is the additional utility provided by one additional unit of consumption. This equation for marginal utility is:

$$\text{MU} = \frac{\text{change in total utility}}{\text{change in quantity}}$$

Notice that marginal utility diminishes as additional units are consumed, which means that each subsequent unit of a good consumed provides less *additional* utility. For example, the first T-shirt José picks is his favorite and it gives him an addition of 22 utils. The fourth T-shirt is just something to wear when all his other clothes are in the wash and yields only 18 additional utils. This is an example of the law of **diminishing marginal utility**, which holds that the additional utility decreases with each unit added. Diminishing marginal utility is another example of the more general law of diminishing returns we learned earlier in the chapter on **Choice in a World of Scarcity**.

The rest of **Table 6.2** shows the quantity of movies that José attends, and his total and marginal utility from seeing each movie. Total utility follows the expected pattern: it increases as the number of movies that José watches rises. Marginal utility also follows the expected pattern: each additional movie brings a smaller gain in utility than the previous one. The first movie José attends is the one he wanted to see the most, and thus provides him with the highest level of utility or satisfaction. The fifth movie he attends is just to kill time. Notice that total utility is also the sum of the marginal utilities. Read the next Work It Out feature for instructions on how to calculate total utility.

T-Shirts (Quantity)	Total Utility	Marginal Utility	Movies (Quantity)	Total Utility	Marginal Utility
1	22	22	**1**	16	16
2	43	21	**2**	31	15
3	63	20	**3**	45	14
4	81	18	**4**	58	13
5	97	16	**5**	70	12
6	111	14	**6**	81	11
7	123	12	**7**	91	10
8	133	10	**8**	100	9

Table 6.2 Total and Marginal Utility

Table 6.3 looks at each point on the budget constraint in **Figure 6.2**, and adds up José's total utility for five possible combinations of T-shirts and movies.

Point	T-Shirts	Movies	Total Utility
P	4	0	81 + 0 = 81
Q	3	2	63 + 31 = 94
R	2	4	43 + 58 = 101
S	1	6	22 + 81 = 103
T	0	8	0 + 100 = 100

Table 6.3 Finding the Choice with the Highest Utility

Work It Out

Calculating Total Utility

Let's look at how José makes his decision in more detail.

Step 1. Observe that, at point Q (for example), José consumes three T-shirts and two movies.

Step 2. Look at **Table 6.2**. You can see from the fourth row/second column that three T-shirts are worth 63 utils. Similarly, the second row/fifth column shows that two movies are worth 31 utils.

Step 3. From this information, you can calculate that point Q has a total utility of 94 (63 + 31).

Step 4. You can repeat the same calculations for each point on **Table 6.3**, in which the total utility numbers are shown in the last column.

For José, the highest total utility for all possible combinations of goods occurs at point S, with a total utility of 103 from consuming one T-shirt and six movies.

Choosing with Marginal Utility

Most people approach their utility-maximizing combination of choices in a step-by-step way. This approach is based on looking at the tradeoffs, measured in terms of marginal utility, of consuming less of one good and more of another.

For example, say that José starts off thinking about spending all his money on T-shirts and choosing point P, which corresponds to four T-shirts and no movies, as **Figure 6.2** illustrates. José chooses this starting point randomly as he has to start somewhere. Then he considers giving up the last T-shirt, the one that provides him the least marginal utility, and using the money he saves to buy two movies instead. **Table 6.4** tracks the step-by-step series of decisions José needs to make (*Key*: T-shirts are $14, movies are $7, and income is $56). The following Work It Out feature explains how marginal utility can effect decision making.

Try	Which Has	Total Utility	Marginal Gain and Loss of Utility, Compared with Previous Choice	Conclusion
Choice 1: P	4 T-shirts and 0 movies	81 from 4 T-shirts + 0 from 0 movies = 81	–	–
Choice 2: Q	3 T-shirts and 2 movies	63 from 3 T-shirts + 31 from 0 movies = 94	Loss of 18 from 1 less T-shirt, but gain of 31 from 2 more movies, for a net utility gain of 13	Q is preferred over P
Choice 3: R	2 T-shirts and 4 movies	43 from 2 T-shirts + 58 from 4 movies = 101	Loss of 20 from 1 less T-shirt, but gain of 27 from two more movies for a net utility gain of 7	R is preferred over Q
Choice 4: S	1 T-shirt and 6 movies	22 from 1 T-shirt + 81 from 6 movies = 103	Loss of 21 from 1 less T-shirt, but gain of 23 from two more movies, for a net utility gain of 2	S is preferred over R

Table 6.4 A Step-by-Step Approach to Maximizing Utility

Try	Which Has	Total Utility	Marginal Gain and Loss of Utility, Compared with Previous Choice	Conclusion
Choice 5: T	0 T-shirts and 8 movies	0 from 0 T-shirts + 100 from 8 movies = 100	Loss of 22 from 1 less T-shirt, but gain of 19 from two more movies, for a net utility loss of 3	S is preferred over T

Table 6.4 A Step-by-Step Approach to Maximizing Utility

Work It Out

Decision Making by Comparing Marginal Utility

José could use the following thought process (if he thought in utils) to make his decision regarding how many T-shirts and movies to purchase:

Step 1. From Table 6.2, José can see that the marginal utility of the fourth T-shirt is 18. If José gives up the fourth T-shirt, then he loses 18 utils.

Step 2. Giving up the fourth T-shirt, however, frees up $14 (the price of a T-shirt), allowing José to buy the first two movies (at $7 each).

Step 3. José knows that the marginal utility of the first movie is 16 and the marginal utility of the second movie is 15. Thus, if José moves from point P to point Q, he gives up 18 utils (from the T-shirt), but gains 31 utils (from the movies).

Step 4. Gaining 31 utils and losing 18 utils is a net gain of 13. This is just another way of saying that the total utility at Q (94 according to the last column in Table 6.3) is 13 more than the total utility at P (81).

Step 5. Thus, for José, it makes sense to give up the fourth T-shirt in order to buy two movies.

José clearly prefers point Q to point P. Now repeat this step-by-step process of decision making with marginal utilities. José thinks about giving up the third T-shirt and surrendering a marginal utility of 20, in exchange for purchasing two more movies that promise a combined marginal utility of 27. José prefers point R to point Q. What if José thinks about going beyond R to point S? Giving up the second T-shirt means a marginal utility loss of 21, and the marginal utility gain from the fifth and sixth movies would combine to make a marginal utility gain of 23, so José prefers point S to R.

However, if José seeks to go beyond point S to point T, he finds that the loss of marginal utility from giving up the first T-shirt is 22, while the marginal utility gain from the last two movies is only a total of 19. If José were to choose point T, his utility would fall to 100. Through these stages of thinking about marginal tradeoffs, José again concludes that S, with one T-shirt and six movies, is the choice that will provide him with the highest level of total utility. This step-by-step approach will reach the same conclusion regardless of José's starting point.

We can develop a more systematic way of using this approach by focusing on satisfaction per dollar. If an item costing $5 yields 10 utils, then it's worth 2 utils per dollar spent. **Marginal utility per dollar** is the amount of additional utility José receives divided by the product's price. Table 6.5 shows the marginal utility per dollar for José's T shirts and movies.

$$\text{marginal utility per dollar} = \frac{\text{marginal utility}}{\text{price}}$$

If José wants to maximize the utility he gets from his limited budget, he will always purchase the item with the greatest marginal utility per dollar of expenditure (assuming he can afford it with his remaining budget). José starts with no purchases. If he purchases a T-shirt, the marginal utility per dollar spent will be 1.6. If he purchases a movie, the marginal utility per dollar spent will be 2.3. Therefore, José's first purchase will be the movie. Why? Because it gives him the highest marginal utility per dollar and is affordable. Next, José will purchase another movie. Why?

Because the marginal utility of the next movie (2.14) is greater than the marginal utility of the next T-shirt (1.6). Note that when José has no T- shirts, the next one is the first one. José will continue to purchase the next good with the highest marginal utility per dollar until he exhausts his budget. He will continue purchasing movies because they give him a greater "bang for the buck" until the sixth movie which gives the same marginal utility per dollar as the first T-shirt purchase. José has just enough budget to purchase both. So in total, José will purchase six movies and one T-shirt.

Quantity of T-Shirts	Total Utility	Marginal Utility	Marginal Utility per Dollar	Quantity of Movies	Total Utility	Marginal Utility	Marginal Utility per Dollar
1	**22**	**22**	**22/$14=1.6**	1	16	16	16/$7=2.3
2	43	21	21/$14=1.5	2	31	15	15/$7=2.14
3	63	20	20/$14=1.4	3	45	14	14/$7=2
4	81	18	18/$14=1.3	4	58	13	13/$7=1.9
5	97	16	16/$14=1.1	5	70	12	12/$7=1.7
6	111	14	14/$14=1	**6**	**81**	**11**	**11/$7=1.6**
7	123	12	12/$14=1.2	7	91	10	10/$7=1.4

Table 6.5 Marginal Utility per Dollar

A Rule for Maximizing Utility

This process of decision making suggests a rule to follow when maximizing utility. Since the price of T-shirts is twice as high as the price of movies, to maximize utility the last T-shirt that José chose needs to provide exactly twice the marginal utility (MU) of the last movie. If the last T-shirt provides less than twice the marginal utility of the last movie, then the T-shirt is providing less “bang for the buck” (i.e., marginal utility per dollar spent) than José would receive from spending the same money on movies. If this is so, José should trade the T-shirt for more movies to increase his total utility.

If the last T-shirt provides more than twice the marginal utility of the last movie, then the T-shirt is providing more “bang for the buck” or marginal utility per dollar, than if the money were spent on movies. As a result, José should buy more T-shirts. Notice that at José’s optimal choice of point S, the marginal utility from the first T-shirt, of 22 is exactly twice the marginal utility of the sixth movie, which is 11. At this choice, the marginal utility per dollar is the same for both goods. This is a tell-tale signal that José has found the point with highest total utility.

We can write this argument as a general rule: If you always choose the item with the greatest marginal utility per dollar spent, when your budget is exhausted, the utility maximizing choice should occur where the marginal utility per dollar spent is the same for both goods.

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

A sensible economizer will pay twice as much for something only if, in the marginal comparison, the item confers twice as much utility. Notice that the formula for the table above is:

$$\frac{22}{\$14} = \frac{11}{\$7}$$
$$1.6 = 1.6$$

The following Work It Out feature provides step by step guidance for this concept of utility-maximizing choices.

Work It Out

Maximizing Utility

The general rule, $\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$, means that the last dollar spent on each good provides exactly the same marginal utility. This is the case at point S. So:

Step 1. If we traded a dollar more of movies for a dollar more of T-shirts, the marginal utility gained from T-shirts would exactly offset the marginal utility lost from fewer movies. In other words, the net gain would be zero.

Step 2. Products, however, usually cost more than a dollar, so we cannot trade a dollar's worth of movies. The best we can do is trade two movies for another T-shirt, since in this example T-shirts cost twice what a movie does.

Step 3. If we trade two movies for one T-shirt, we would end up at point R (two T-shirts and four movies).

Step 4. Choice 4 in **Table 6.4** shows that if we move to point R, we would gain 21 utils from one more T-shirt, but lose 23 utils from two fewer movies, so we would end up with less total utility at point R.

In short, the general rule shows us the utility-maximizing choice, which is called the **consumer equilibrium**.

There is another equivalent way to think about this. We can also express the general rule as *the ratio of the prices of the two goods should be equal to the ratio of the marginal utilities.* When we divide the price of good 1 by the price of good 2, at the utility-maximizing point this will equal the marginal utility of good 1 divided by the marginal utility of good 2.

$$\frac{P_1}{P_2} = \frac{MU_1}{MU_2}$$

Along the budget constraint, the total price of the two goods remains the same, so the ratio of the prices does not change. However, the marginal utility of the two goods changes with the quantities consumed. At the optimal choice of one T-shirt and six movies, point S, the ratio of marginal utility to price for T-shirts (22:14) matches the ratio of marginal utility to price for movies (of 11:7).

Measuring Utility with Numbers

This discussion of utility began with an assumption that it is possible to place numerical values on utility, an assumption that may seem questionable. You can buy a thermometer for measuring temperature at the hardware store, but what store sells an "utilimometer" for measuring utility? While measuring utility with numbers is a convenient assumption to clarify the explanation, the key assumption is not that an outside party can measure utility but only that individuals can decide which of two alternatives they prefer.

To understand this point, think back to the step-by-step process of finding the choice with highest total utility by comparing the marginal utility you gain and lose from different choices along the budget constraint. As José compares each choice along his budget constraint to the previous choice, what matters is not the specific numbers that he places on his utility—or whether he uses any numbers at all—but only that he personally can identify which choices he prefers.

In this way, the step-by-step process of choosing the highest level of utility resembles rather closely how many people make consumption decisions. We think about what will make us the happiest. We think about what things cost. We think about buying a little more of one item and giving up a little of something else. We choose what provides us with the greatest level of satisfaction. The vocabulary of comparing the points along a budget constraint and total and marginal utility is just a set of tools for discussing this everyday process in a clear and specific manner. It is welcome news that specific utility numbers are not central to the argument, since a good utilimometer is hard to find. Do not worry—while we cannot measure utils, by the end of the next module, we will have transformed our analysis into something we can measure—demand.

6.2 | How Changes in Income and Prices Affect Consumption Choices

By the end of this section, you will be able to:

- Explain how income, prices, and preferences affect consumer choices
- Contrast the substitution effect and the income effect
- Utilize concepts of demand to analyze consumer choices
- Apply utility-maximizing choices to governments and businesses

Just as we can use utility and marginal utility to discuss making consumer choices along a budget constraint, we can also use these ideas to think about how consumer choices change when the budget constraint shifts in response to changes in income or price. Because we can use the budget constraint framework to analyze how quantities demanded change because of price movements, the budget constraint model can illustrate the underlying logic behind demand curves.

How Changes in Income Affect Consumer Choices

Let's begin with a concrete example illustrating how changes in income level affect consumer choices. **Figure 6.3** shows a budget constraint that represents Kimberly's choice between concert tickets at $50 each and getting away overnight to a bed-and-breakfast for $200 per night. Kimberly has $1,000 per year to spend between these two choices. After thinking about her total utility and marginal utility and applying the decision rule that the ratio of the marginal utilities to the prices should be equal between the two products, Kimberly chooses point M, with eight concerts and three overnight getaways as her utility-maximizing choice.

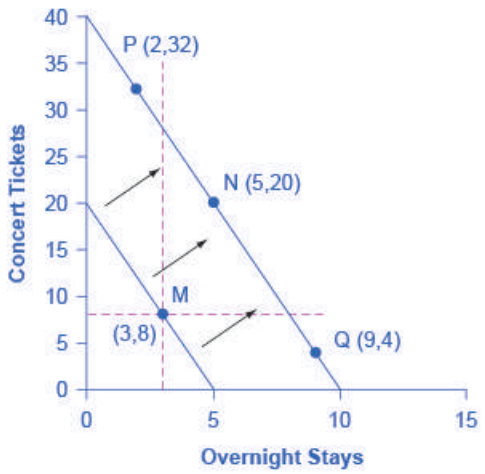


Figure 6.3 How a Change in Income Affects Consumption Choices The utility-maximizing choice on the original budget constraint is M. The dashed horizontal and vertical lines extending through point M allow you to see at a glance whether the quantity consumed of goods on the new budget constraint is higher or lower than on the original budget constraint. On the new budget constraint, Kimberly will make a choice like N if both goods are normal goods. If overnight stays is an inferior good, Kimberly will make a choice like P. If concert tickets are an inferior good, Kimberly will make a choice like Q.

Now, assume that the income Kimberly has to spend on these two items rises to $2,000 per year, causing her budget constraint to shift out to the right. How does this rise in income alter her utility-maximizing choice? Kimberly will again consider the utility and marginal utility that she receives from concert tickets and overnight getaways and seek her utility-maximizing choice on the new budget line, but how will her new choice relate to her original choice?

We can replace the possible choices along the new budget constraint into three groups, which the dashed horizontal and vertical lines that pass through the original choice M in the figure divide. All choices on the upper left of the new budget constraint that are to the left of the vertical dashed line, like choice P with two overnight stays and 32

concert tickets, involve less of the good on the horizontal axis but much more of the good on the vertical axis. All choices to the right of the vertical dashed line and above the horizontal dashed line—like choice N with five overnight getaways and 20 concert tickets—have more consumption of both goods. Finally, all choices that are to the right of the vertical dashed line but below the horizontal dashed line, like choice Q with four concerts and nine overnight getaways, involve less of the good on the vertical axis but much more of the good on the horizontal axis.

All of these choices are theoretically possible, depending on Kimberly's personal preferences as expressed through the total and marginal utility she would receive from consuming these two goods. When income rises, the most common reaction is to purchase more of both goods, like choice N, which is to the upper right relative to Kimberly's original choice M, although exactly how much more of each good will vary according to personal taste. Conversely, when income falls, the most typical reaction is to purchase less of both goods. As we defined in the chapter on **Demand and Supply** and again in the chapter on **Elasticity**, we call goods and services normal goods when a rise in income leads to a rise in the quantity consumed of that good and a fall in income leads to a fall in quantity consumed.

However, depending on Kimberly's preferences, a rise in income could cause consumption of one good to increase while consumption of the other good declines. A choice like P means that a rise in income caused her quantity consumed of overnight stays to decline, while a choice like Q would mean that a rise in income caused her quantity of concerts to decline. Goods where demand declines as income rises (or conversely, where the demand rises as income falls) are called "inferior goods." An inferior good occurs when people trim back on a good as income rises, because they can now afford the more expensive choices that they prefer. For example, a higher-income household might eat fewer hamburgers or be less likely to buy a used car, and instead eat more steak and buy a new car.

How Price Changes Affect Consumer Choices

For analyzing the possible effect of a change in price on consumption, let's again use a concrete example. **Figure 6.4** represents Sergei's consumer choice, who chooses between purchasing baseball bats and cameras. A price increase for baseball bats would have no effect on the ability to purchase cameras, but it would reduce the number of bats Sergei could afford to buy. Thus a price increase for baseball bats, the good on the horizontal axis, causes the budget constraint to rotate inward, as if on a hinge, from the vertical axis. As in the previous section, the point labeled M represents the originally preferred point on the original budget constraint, which Sergei has chosen after contemplating his total utility and marginal utility and the tradeoffs involved along the budget constraint. In this example, the units along the horizontal and vertical axes are not numbered, so the discussion must focus on whether Sergei will consume more or less of certain goods, not on numerical amounts.

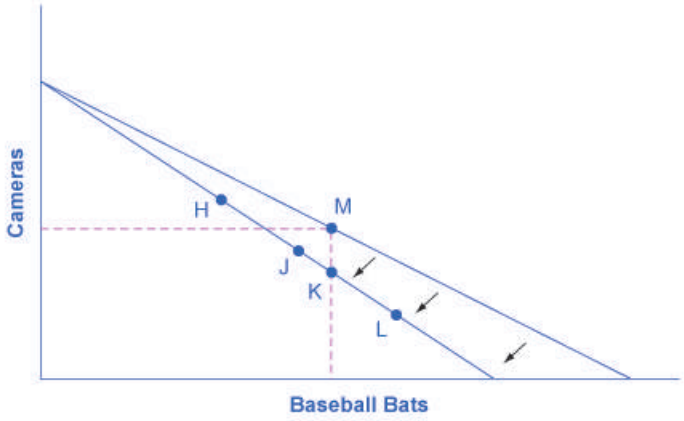


Figure 6.4 How a Change in Price Affects Consumption Choices The original utility-maximizing choice is M. When the price rises, the budget constraint rotates clockwise. The dashed lines make it possible to see at a glance whether the new consumption choice involves less of both goods, or less of one good and more of the other. The new possible choices would be fewer baseball bats and more cameras, like point H, or less of both goods, as at point J. Choice K would mean that the higher price of bats led to exactly the same quantity of bat consumption, but fewer cameras. Theoretically possible, but unlikely in the real world, we rule out choices like L because they would mean that a higher price for baseball bats means a greater consumption of baseball bats.

After the price increase, Sergei will make a choice along the new budget constraint. Again, we can divide his choices into three segments by the dashed vertical and horizontal lines. In the upper left portion of the new budget constraint,

at a choice like H, Sergei consumes more cameras and fewer bats. In the central portion of the new budget constraint, at a choice like J, he consumes less of both goods. At the right-hand end, at a choice like L, he consumes more bats but fewer cameras.

The typical response to higher prices is that a person chooses to consume less of the product with the higher price. This occurs for two reasons, and both effects can occur simultaneously. The **substitution effect** occurs when a price changes and consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price. The **income effect** is that a higher price means, in effect, the buying power of income has been reduced (even though actual income has not changed), which leads to buying less of the good (when the good is normal). In this example, the higher price for baseball bats would cause Sergei to buy fewer bats for both reasons. Exactly how much will a higher price for bats cause Sergei's bat consumption to fall? **Figure 6.4** suggests a range of possibilities. Sergei might react to a higher price for baseball bats by purchasing the same quantity of bats, but cutting his camera consumption. This choice is the point K on the new budget constraint, straight below the original choice M. Alternatively, Sergei might react by dramatically reducing his bat purchases and instead buy more cameras.

The key is that it would be imprudent to assume that a change in baseball bats will only or primarily affect the good's price whose price is changed, while the quantity consumed of other goods remains the same. Since Sergei purchases all his products out of the same budget, a change in the price of one good can also have a range of effects, either positive or negative, on the quantity consumed of other goods.

In short, a higher price typically causes reduced consumption of the good in question, but it can affect the consumption of other goods as well.

Link It Up

Read this article (http://openstaxcollege.org/l/vending) about the potential of variable prices in vending machines.

The Foundations of Demand Curves

Changes in the price of a good lead the budget constraint to rotate. A rotation in the budget constraint means that when individuals are seeking their highest utility, the quantity that is demanded of that good will change. In this way, the logical foundations of demand curves—which show a connection between prices and quantity demanded—are based on the underlying idea of individuals seeking utility. **Figure 6.5** (a) shows a budget constraint with a choice between housing and “everything else.” (Putting “everything else” on the vertical axis can be a useful approach in some cases, especially when the focus of the analysis is on one particular good.) We label the preferred choice on the original budget constraint that provides the highest possible utility M_0. The other three budget constraints represent successively higher prices for housing of P_1, P_2, and P_3. As the budget constraint rotates in, and in, and in again, we label the utility-maximizing choices M_1, M_2, and M_3, and the quantity demanded of housing falls from Q_0 to Q_1 to Q_2 to Q_3.

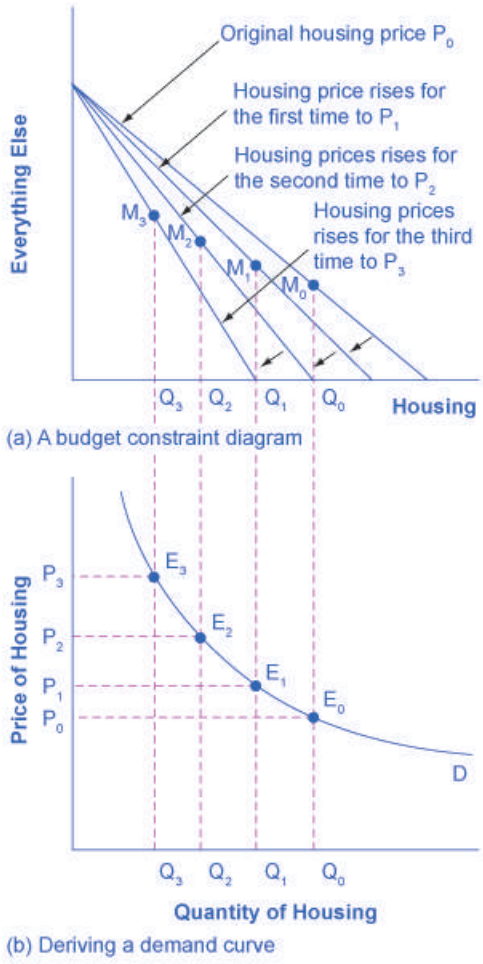


Figure 6.5 The Foundations of a Demand Curve: An Example of Housing (a) As the price increases from P_0 to P_1 to P_2 to P_3, the budget constraint on the upper part of the diagram rotates clockwise. The utility-maximizing choice changes from M_0 to M_1 to M_2 to M_3. As a result, the quantity demanded of housing shifts from Q_0 to Q_1 to Q_2 to Q_3, *ceteris paribus*. (b) The demand curve graphs each combination of the price of housing and the quantity of housing demanded, *ceteris paribus*. The quantities of housing are the same at the points on both (a) and (b). Thus, the original price of housing (P_0) and the original quantity of housing (Q_0) appear on the demand curve as point E_0. The higher price of housing (P_1) and the corresponding lower quantity demanded of housing (Q_1) appear on the demand curve as point E_1.

Thus, as the price of housing rises, the budget constraint rotates clockwise and the quantity consumed of housing falls, *ceteris paribus* (meaning, with all other things being the same). We graph this relationship—the price of housing rising from P_0 to P_1 to P_2 to P_3, while the quantity of housing demanded falls from Q_0 to Q_1 to Q_2 to Q_3—on the demand curve in **Figure 6.5** (b). The vertical dashed lines stretching between the top and bottom of **Figure 6.5** show that the quantity of housing demanded at each point is the same in both (a) and (b). We ultimately determine the shape of a demand curve by the underlying choices about maximizing utility subject to a budget constraint. While economists may not be able to measure "utils," they can certainly measure price and quantity demanded.

Applications in Government and Business

The budget constraint framework for making utility-maximizing choices offers a reminder that people can react to a change in price or income in a range of different ways. For example, in the winter months of 2005, costs for heating homes increased significantly in many parts of the country as prices for natural gas and electricity soared, due in large part to the disruption caused by Hurricanes Katrina and Rita. Some people reacted by reducing the quantity demanded of energy; for example, by turning down the thermostats in their homes by a few degrees and wearing a heavier sweater inside. Even so, many home heating bills rose, so people adjusted their consumption in other ways, too. As you learned in the chapter on **Elasticity**, the short run demand for home heating is generally inelastic. Each

household cut back on what it valued least on the margin. For some it might have been some dinners out, or a vacation, or postponing buying a new refrigerator or a new car. Sharply higher energy prices can have effects beyond the energy market, leading to a widespread reduction in purchasing throughout the rest of the economy.

A similar issue arises when the government imposes taxes on certain products, such as on gasoline, cigarettes, and alcohol. Say that a tax on alcohol leads to a higher price at the liquor store. The higher price of alcohol causes the budget constraint to pivot left, and alcoholic beverage consumption is likely to decrease. However, people may also react to the higher price of alcoholic beverages by cutting back on other purchases. For example, they might cut back on snacks at restaurants like chicken wings and nachos. It would be unwise to assume that the liquor industry is the only one affected by the tax on alcoholic beverages. Read the next Clear It Up to learn about how who controls the household income influences buying decisions.

The Unifying Power of the Utility-Maximizing Budget Set Framework

An interaction between prices, budget constraints, and personal preferences determine household choices. The flexible and powerful terminology of utility-maximizing gives economists a vocabulary for bringing these elements together.

Not even economists believe that people walk around mumbling about their marginal utilities before they walk into a shopping mall, accept a job, or make a deposit in a savings account. However, economists do believe that individuals seek their own satisfaction or utility and that people often decide to try a little less of one thing and a little more of another. If we accept these assumptions, then the idea of utility-maximizing households facing budget constraints becomes highly plausible.

Does who controls household income make a difference?

In the mid-1970s, the United Kingdom made an interesting policy change in its “child allowance” policy. This program provides a fixed amount of money per child to every family, regardless of family income. Traditionally, the child allowance had been distributed to families by withholding less in taxes from the paycheck of the family wage earner—typically the father in this time period. The new policy instead provided the child allowance as a cash payment to the mother. As a result of this change, households have the same level of income and face the same prices in the market, but the money is more likely to be in the mother's purse than in the father's wallet.

Should this change in policy alter household consumption patterns? Basic models of consumption decisions, of the sort that we examined in this chapter, assume that it does not matter whether the mother or the father receives the money, because both parents seek to maximize the family's utility as a whole. In effect, this model assumes that everyone in the family has the same preferences.

In reality, the share of that the father or mother controls does affect what the household consumes. When the mother controls a larger share of family income a number of studies, in the United Kingdom and in a wide variety of other countries, have found that the family tends to spend more on restaurant meals, child care, and women’s clothing, and less on alcohol and tobacco. As the mother controls a larger share of household resources, children’s health improves, too. These findings suggest that when providing assistance to poor families, in high-income countries and low-income countries alike, the monetary amount of assistance is not all that matters: it also matters which family member actually receives the money.

The budget constraint framework serves as a constant reminder to think about the full range of effects that can arise from changes in income or price, not just effects on the one product that might seem most immediately affected.

6.3 | Labor-Leisure Choices

By the end of this section, you will be able to:

- Interpret labor-leisure budget constraint graphs
- Predict consumer choices based on wages and other compensation
- Explain the backward-bending supply curve of labor

People do not obtain utility just from products they purchase. They also obtain utility from leisure time. Leisure time is time not spent at work. The decision-making process of a utility-maximizing household applies to what quantity of hours to work in much the same way that it applies to purchases of goods and services. Choices made along the labor-leisure budget constraint, as wages shift, provide the logical underpinning for the labor supply curve. The discussion also offers some insights about the range of possible reactions when people receive higher wages, and specifically about the claim that if people are paid higher wages, they will work a greater quantity of hours—assuming that they have a say in the matter.

According to the Bureau of Labor Statistics, U.S. workers averaged 38.6 hours per week on the job in 2014. This average includes part-time workers; for full-time workers only, the average was 42.5 hours per week. **Table 6.6** shows that more than half of all workers are on the job 35 to 48 hours per week, but significant proportions work more or less than this amount.

Table 6.7 breaks down the average hourly compensation received by private industry workers, including wages and benefits. Wages and salaries are about three-quarters of total compensation received by workers; the rest is in the form of health insurance, vacation pay, and other benefits. The compensation workers receive differs for many reasons, including experience, education, skill, talent, membership in a labor union, and the presence of discrimination against certain groups in the labor market. Issues surrounding the inequality of incomes in a market-oriented economy are explored in the chapters on **Poverty and Economic Inequality (http://cnx.org/content/m57266/latest/)** and **Issue in Labor Markets: Unions, Discrimination, Immigration (http://cnx.org/content/m57275/latest/)** .

Hours Worked per Week	Number of Workers	Percentage of Workforce
1–14 hours	6.9 million	5.0%
15–34 hours	27.6 million	20.1%
35–40 hours	68.5 million	49.9%
41–48 hours	11.9 million	8.6%
49–59 hours	13.3 million	9.6%
60 hours and over	9.3 million	6.8%

Table 6.6 Persons at Work, by Average Hours Worked per Week in 2013 (Total number of workers: 137.7 million) (Source: http://www.bls.gov/news.release/empsit.t18.htm)

Compensation, Wage, Salary, and Benefits	$30.92 per hour
Wages and Salaries	$20.92
Benefits	

Table 6.7 Hourly Compensation: Wages, Benefits, and Taxes in 2014 (Source: http://www.bls.gov/news.release/pdf/ecec.pdf)

Compensation, Wage, Salary, and Benefits	$30.92 per hour
Vacation	$2.09
Supplemental Pay	$0.84
Insurance	$2.15
Health Benefits	$2.36
Retirement and Savings	$1.24
Defined Benefit	$0.57
Defined Contribution	$0.064
Legally Required	$2.46

Table 6.7 Hourly Compensation: Wages, Benefits, and Taxes in 2014 (Source: http://www.bls.gov/news.release/pdf/ecec.pdf)

The Labor-Leisure Budget Constraint

How do workers make decisions about the number of hours to work? Again, let's proceed with a concrete example. The economic logic is precisely the same as in the case of a consumption choice budget constraint, but the labels are different on a labor-leisure budget constraint.

Vivian has 70 hours per week that she could devote either to work or to leisure, and her wage is $10/hour. The lower budget constraint in **Figure 6.6** shows Vivian's possible choices. The horizontal axis of this diagram measures both leisure and labor, by showing how Vivian's time is divided between leisure and labor. Hours of leisure are measured from left to right on the horizontal axis, while hours of labor are measured from right to left. Vivian will compare choices along this budget constraint, ranging from 70 hours of leisure and no income at point S to zero hours of leisure and $700 of income at point L. She will choose the point that provides her with the highest total utility. For this example, let's assume that Vivian's utility-maximizing choice occurs at O, with 30 hours of leisure, 40 hours of work, and $400 in weekly income.

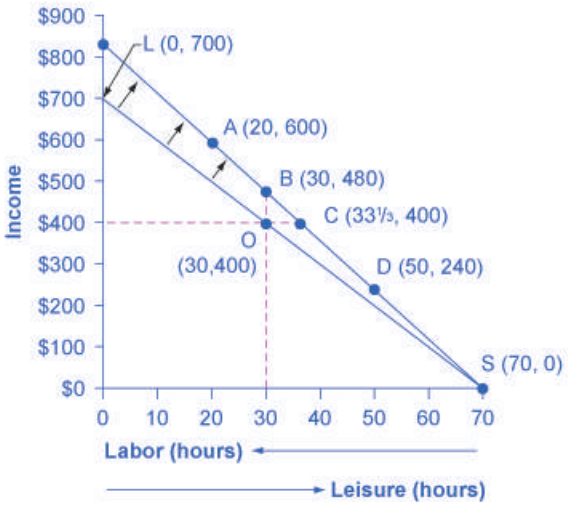


Figure 6.6 How a Rise in Wages Alters the Utility-Maximizing Choice Vivian's original choice is point O on the lower opportunity set. A rise in her wage causes her opportunity set to swing upward. In response to the increase in wages, Vivian can make a range of different choices available to her: a choice like D, which involves less work; and a choice like B, which involves the same amount of work but more income; or a choice like A, which involves more work and considerably more income. Vivian's personal preferences will determine which choice she makes.

For Vivian to discover the labor-leisure choice that will maximize her utility, she does not have to place numerical values on the total and marginal utility that she would receive from every level of income and leisure. All that really matters is that Vivian can compare, in her own mind, whether she would prefer more leisure or more income, given the tradeoffs she faces. If Vivian can say to herself: "I'd really rather work a little less and have more leisure, even if it means less income," or "I'd be willing to work more hours to make some extra income," then as she gradually moves in the direction of her preferences, she will seek out the utility-maximizing choice on her labor-leisure budget constraint.

Now imagine that Vivian's wage level increases to $12/hour. A higher wage will mean a new budget constraint that tilts up more steeply; conversely, a lower wage would have led to a new budget constraint that was flatter. How will a change in the wage and the corresponding shift in the budget constraint affect Vivian's decisions about how many hours to work?

Vivian's choices of quantity of hours to work and income along her new budget constraint can be divided into several categories, using the dashed horizontal and vertical lines in **Figure 6.6** that go through her original choice (O). One set of choices in the upper-left portion of the new budget constraint involves more hours of work (that is, less leisure) and more income, at a point like A with 20 hours of leisure, 50 hours of work, and $600 of income (that is, 50 hours of work multiplied by the new wage of $12 per hour). A second choice would be to work exactly the same 40 hours, and to take the benefits of the higher wage in the form of income that would now be $480, at choice B. A third choice would involve more leisure and the same income at point C (that is, 33-1/3 hours of work multiplied by the new wage of $12 per hour equals $400 of total income). A fourth choice would involve less income and much more leisure at a point like D, with a choice like 50 hours of leisure, 20 hours of work, and $240 in income.

In effect, Vivian can choose whether to receive the benefits of her wage increase in the form of more income, or more leisure, or some mixture of these two. With this range of possibilities, it would be unwise to assume that Vivian (or anyone else) will necessarily react to a wage increase by working substantially more hours. Maybe they will; maybe they will not.

Applications of Utility Maximizing with the Labor-Leisure Budget Constraint

The theoretical insight that higher wages will sometimes cause an increase in hours worked, sometimes cause hours worked not to change by much, and sometimes cause hours worked to decline, has led to labor supply curves that look like the one in **Figure 6.7**. The bottom-left portion of the labor supply curve slopes upward, which reflects the situation of a person who reacts to a higher wage by supplying a greater quantity of labor. The middle, close-to-vertical portion of the labor supply curve reflects the situation of a person who reacts to a higher wage by supplying about the same quantity of labor. The very top portion of the labor supply curve is called a **backward-bending supply curve for labor**, which is the situation of high-wage people who can earn so much that they respond to a still-higher wage by working fewer hours. Read the following Clear It Up feature for more on the number of hours the average person works each year.

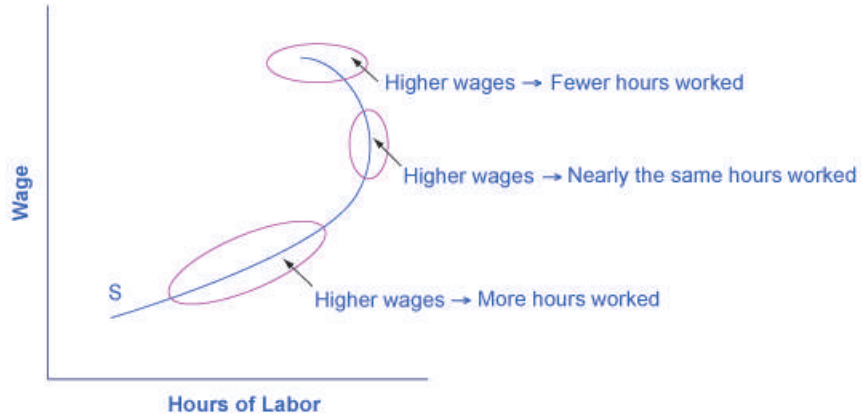


Figure 6.7 A Backward-Bending Supply Curve of Labor The bottom upward-sloping portion of the labor supply curve shows that as wages increase over this range, the quantity of hours worked also increases. The middle, nearly vertical portion of the labor supply curve shows that as wages increase over this range, the quantity of hours worked changes very little. The backward-bending portion of the labor supply curve at the top shows that as wages increase over this range, the quantity of hours worked actually decreases. All three of these possibilities can be derived from how a change in wages causes movement in the labor-leisure budget constraint, and thus different choices by individuals.

Clear It Up

Is America a nation of workaholics?

Americans work a lot. **Table 6.8** shows average hours worked per year in the United States, Canada, Japan, and several European countries, with data from 2013. To get a perspective on these numbers, someone who works 40 hours per week for 50 weeks per year, with two weeks off, would work 2,000 hours per year. The gap in hours worked is a little astonishing; the 250 to 300 hour gap between how much Americans work and how much Germans or the French work amounts to roughly six to seven weeks less of work per year. Economists who study these international patterns debate the extent to which average Americans and Japanese have a preference for working more than, say, Germans, or whether German workers and employers face particular kinds of taxes and regulations that lead to fewer hours worked. Many countries have laws that regulate the work week and dictate holidays and the standards of "normal" vacation time vary from country to country. It is also interesting to take the amount of time spent working in context; it is estimated that in the late nineteenth century in the United States, the average work week was over 60 hours per week—leaving little to no time for leisure.

Country	Average Annual Hours Actually Worked per Employed Person
United States	1,824
Spain	1,799
Japan	1,759
Canada	1,751
United Kingdom	1,669
Sweden	1,585
Germany	1,443
France	1,441

Table 6.8 Average Hours Worked Per Year in Select Countries (Source: http://stats.oecd.org/Index.aspx?DataSetCode=ANHRS)

The different responses to a rise in wages—more hours worked, the same hours worked, or fewer hours worked—are patterns exhibited by different groups of workers in the U.S. economy. Many full-time workers have jobs where the number of hours is held relatively fixed, partly by their own choice and partly by their employer's practices. These workers do not much change their hours worked as wages rise or fall, so their supply curve of labor is inelastic. However, part-time workers and younger workers tend to be more flexible in their hours, and more ready to increase hours worked when wages are high or cut back when wages fall.

The backward-bending supply curve for labor, when workers react to higher wages by working fewer hours and having more income, is not observed often in the short run. However, some well-paid professionals, like dentists or accountants, may react to higher wages by choosing to limit the number of hours, perhaps by taking especially long vacations, or taking every other Friday off. Over a long-term perspective, the backward-bending supply curve for labor is common. Over the last century, Americans have reacted to gradually rising wages by working fewer hours; for example, the length of the average work-week has fallen from about 60 hours per week in 1900 to the present average of less than 40 hours per week.

Recognizing that workers have a range of possible reactions to a change in wages casts some fresh insight on a perennial political debate: the claim that a reduction in income taxes—which would, in effect, allow people to earn

more per hour—will encourage people to work more. The leisure-income budget set points out that this connection will not hold true for all workers. Some people, especially part-timers, may react to higher wages by working more. Many will work the same number of hours. Some people, especially those whose incomes are already high, may react to the tax cut by working *fewer* hours. Of course, cutting taxes may be a good or a bad idea for a variety of reasons, not just because of its impact on work incentives, but the specific claim that tax cuts will lead people to work more hours is only likely to hold for specific groups of workers and will depend on how and for whom taxes are cut.

6.4 | Intertemporal Choices in Financial Capital Markets

By the end of this section, you will be able to:

- Evaluate the reasons for making intertemporal choices
- Interpret an intertemporal budget constraint
- Analyze why people in America tend to save such a small percentage of their income

Rates of saving in America have never been especially high, but they seem to have dipped even lower in recent years, as the data from the Bureau of Economic Analysis in **Figure 6.8** show. A decision about how much to save can be represented using an intertemporal budget constraint. Household decisions about the quantity of financial savings show the same underlying pattern of logic as the consumption choice decision and the labor-leisure decision.

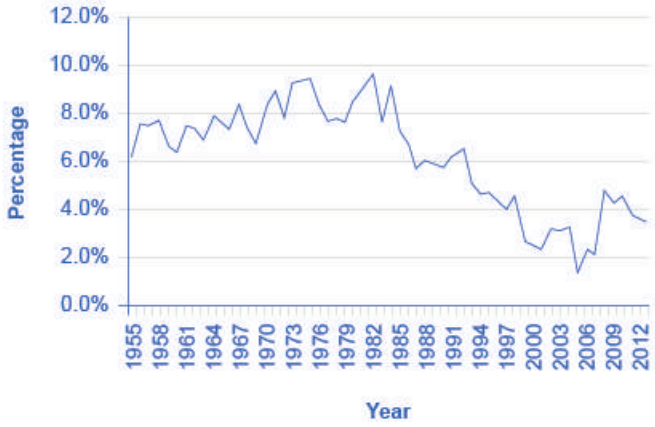


Figure 6.8 Personal Savings as a Percentage of Personal Income Personal savings were about 7 to 11% of personal income for most of the years from the late 1950s up to the early 1990s. Since then, the rate of personal savings has fallen substantially, although it seems to have bounced back a bit since 2008. (Source: http://www.bea.gov/newsreleases/national/pi/pinewsrelease.htm)

The discussion of financial saving here will not focus on the specific financial investment choices, like bank accounts, stocks, bonds, mutual funds, or owning a house or gold coins. The characteristics of these specific financial investments, along with the risks and tradeoffs they pose, are detailed in the **Labor and Financial Markets (http://cnx.org/content/m57219/latest/)** chapter. Here, the focus is saving in total—that is, on how a household determines how much to consume in the present and how much to save, given the expected rate of return (or interest rate), and how the quantity of saving alters when the rate of return changes.

Using Marginal Utility to Make Intertemporal Choices

Savings behavior varies considerably across households. One factor is that households with higher incomes tend to save a larger percentage of their income. This pattern makes intuitive sense; a well-to-do family has the flexibility in its budget to save 20–25% of income, while a poor family struggling to keep food on the table will find it harder to put money aside.

Another factor that causes personal saving to vary is personal preferences. Some people may prefer to consume more now, and let the future look after itself. Others may wish to enjoy a lavish retirement, complete with expensive vacations, or to pile up money that they can pass along to their grandchildren. There are savers and spendthrifts among the young, middle-aged, and old, and among those with high, middle, and low income levels.

Consider this example: Yelberton is a young man starting off at his first job. He thinks of the "present" as his working life and the "future" as after retirement. Yelberton's plan is to save money from ages 30 to 60, retire at age 60, and then live off his retirement money from ages 60 to 85. On average, therefore, he will be saving for 30 years. If the rate of return that he can receive is 6% per year, then $1 saved in the present would build up to $5.74 after 30 years (using the formula for compound interest, $\$1(1 + 0.06)^{30} = \$5.74$). Say that Yelberton will earn $1,000,000 over the 30 years from age 30 to age 60 (this amount is approximately an annual salary of $33,333 multiplied by 30 years). The question for Yelberton is how much of those lifetime earnings to consume during his working life, and how much to put aside until after retirement. This example is obviously built on simplifying assumptions, but it does convey the basic life-cycle choice of saving during working life for future consumption after retirement.

Figure 6.9 and **Table 6.9** show Yelberton's intertemporal budget constraint. Yelberton's choice involves comparing the utility of present consumption during his working life and future consumption after retirement. The rate of return that determines the slope of the intertemporal budget line between present consumption and future consumption in this example is the annual interest rate that he would earn on his savings, compounded over the 30 years of his working life. (For simplicity, we are assuming that any savings from current income will compound for 30 years.) Thus, in the lower budget constraint line on the figure, future consumption grows by increments of $574,000, because each time $100,000 is saved in the present, it compounds to $574,000 after 30 years at a 6% interest rate. If some of the numbers on the future consumption axis look bizarrely large, remember that this occurs because of the power of compound interest over substantial periods of time, and because the figure is grouping together all of Yelberton's saving for retirement over his lifetime.

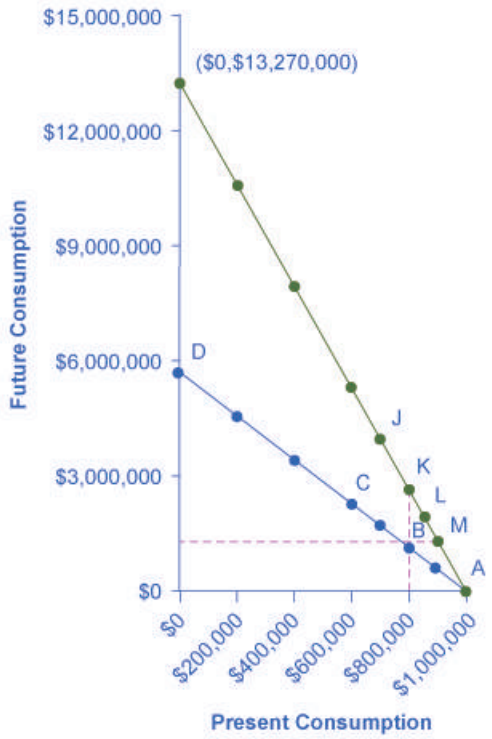


Figure 6.9 Yelberton's Choice: The Intertemporal Budget Set Yelberton will make a choice between present and future consumption. With an annual rate of return of 6%, he decides that his utility will be highest at point B, which represents a choice of $800,000 in present consumption and $1,148,000 in future consumption. When the annual rate of return rises to 9%, the intertemporal budget constraint pivots up. Yelberton could choose to take the gains from this higher rate of return in several forms: more present saving and much higher future consumption (J), the same present saving and higher future consumption (K), more present consumption and more future consumption (L), or more present consumption and the same future consumption (M).

Present Consumption	Present Savings	Future Consumption (6% annual return)	Future Consumption (9% annual return)
$1,000,000	0	0	0
$900,000	$100,000	$574,000	$1,327,000
$800,000	$200,000	$1,148,000	$2,654,000
$700,000	$300,000	$1,722,000	$3,981,000
$600,000	$400,000	$2,296,000	$5,308,000
$400,000	$600,000	$3,444,000	$7,962,000
$200,000	$800,000	$4,592,000	$10,616,000
0	$1,000,000	$5,740,000	$13,270,000

Table 6.9 Yelburton's Intertemporal Budget Constraint

Yelberton will compare the different choices along the budget constraint and choose the one that provides him with the highest utility. For example, he will compare the utility he would receive from a choice like point A, with consumption of $1 million in the present, zero savings, and zero future consumption; point B, with present consumption of $800,000, savings of $200,000, and future consumption of $1,148,000; point C, with present consumption of $600,000, savings of $400,000, and future consumption of $2,296,000; or even choice D, with present consumption of zero, savings of $1,000,000, and future consumption of $5,740,000. Yelberton will also ask himself questions like these: "Would I prefer to consume a little less in the present, save more, and have more future consumption?" or "Would I prefer to consume a little more in the present, save less, and have less future consumption?" By considering marginal changes toward more or less consumption, he can seek out the choice that will provide him with the highest level of utility.

Let us say that Yelberton's preferred choice is B. Imagine that Yelberton's annual rate of return raises from 6% to 9%. In this case, each time he saves $100,000 in the present, it will be worth $1,327,000 in 30 years from now (using the formula for compound interest that $100,000 $(1 + 0.09)^{30}$ = $1,327,000). A change in rate of return alters the slope of the intertemporal budget constraint: a higher rate of return or interest rate will cause the budget line to pivot upward, while a lower rate of return will cause it to pivot downward. If Yelberton were to consume nothing in the present and save all $1,000,000, with a 9% rate of return, his future consumption would be $13,270,000, as shown on **Figure 6.9**.

As the rate of return rises, Yelberton considers a range of choices on the new intertemporal budget constraint. The dashed vertical and horizontal lines running through the original choice B help to illustrate his range of options. One choice is to reduce present consumption (that is, to save more) and to have considerably higher future consumption at a point like J above and to the left of his original choice B. A second choice would be to keep the level of present consumption and savings the same, and to receive the benefits of the higher rate of return entirely in the form of higher future consumption, which would be choice K.

As a third choice Yelberton could have both more present consumption—that is, less savings—but still have higher future consumption because of the higher interest rate, which would be choice like L, above and to the right of his original choice B. Thus, the higher rate of return might cause Yelberton to save more, or less, or the same amount, depending on his own preferences. A fourth choice would be that Yelberton could react to the higher rate of return by increasing his current consumption and leaving his future consumption unchanged, as at point M directly to the right of his original choice B. The actual choice of what quantity to save and how saving will respond to changes in the rate of return will vary from person to person, according to the choice that will maximize each person's utility.

Applications of the Model of Intertemporal Choice

The theoretical model of the intertemporal budget constraint suggests that when the rate of return rises, the quantity of saving may rise, fall, or remain the same, depending on the preferences of individuals. For the U.S. economy as a whole, the most common pattern seems to be that the quantity of savings does not adjust much to changes in the rate of

return. As a practical matter, many households either save at a fairly steady pace, by putting regular contributions into a retirement account or by making regular payments as they buy a house, or they do not save much at all. Of course, some people will have preferences that cause them to react to a higher rate of return by increasing their quantity of saving; others will react to a higher rate of return by noticing that with a higher rate of return, they can save less in the present and still have higher future consumption.

One prominent example in which a higher rate of return leads to a lower savings rate occurs when firms save money because they have promised to pay workers a certain fixed level of pension benefits after retirement. When rates of return rise, those companies can save less money in the present in their pension fund and still have enough to pay the promised retirement benefits in the future.

This insight suggests some skepticism about political proposals to encourage higher savings by providing savers with a higher rate of return. For example, Individual Retirement Accounts (IRAs) and 401(k) accounts are special savings accounts where the money going into the account is not taxed until it is taken out many years later, after retirement. The main difference between these accounts is that an IRA is usually set up by an individual, while a 401(k) needs to be set up through an employer. By not taxing savings in the present, the effect of an IRA or a 401(k) is to increase the return to saving in these accounts.

IRA and 401(k) accounts have attracted a large quantity of savings since they became common in the late 1980s and early 1990s. In fact, the amount of IRAs rose from $239 billion in 1992 to $3.7 trillion in 2005, then to over $5 trillion in 2012, as per the Investment Company Institute, a national association of U.S. investment companies. However, overall U.S. personal savings, as discussed earlier, actually dropped from low to lower in the late 1990s and into the 2000s. Evidently, the larger amounts in these retirement accounts are being offset, in the economy as a whole, either by less savings in other kinds of accounts, or by a larger amount of borrowing (that is, negative savings). The following Clear It Up further explores America's saving rates.

A rise in interest rates makes it easier for people to enjoy higher future consumption. But it also allows them to enjoy higher present consumption, if that is what these individuals desire. Again, a change in prices—in this case, in interest rates—leads to a range of possible outcomes.

Clear It Up

How does America's saving rates compare to other countries?

By international standards, Americans do not save a high proportion of their income, as Table 6.10 shows. The rate of gross national saving includes saving by individuals, businesses, and government. By this measure, U.S. national savings amount to 17% of the size of the U.S. GDP, which measures the size of the U.S. economy. The comparable world average rate of savings is 22%.

Country	Gross Domestic Savings as a Percentage of GDP
China	51%
India	30%
Russia	28%
Mexico	22%
Germany	26%
Japan	22%
Canada	21%
France	21%
Brazil	15%
United States	17%
United Kingdom	13%

Table 6.10 National Savings in Select Countries (Source: http://data.worldbank.org/indicator/NY.GNS.ICTR.ZS)

The Unifying Power of the Utility-Maximizing Budget Set Framework

The choices of households are determined by an interaction between prices, budget constraints, and personal preferences. The flexible and powerful terminology of utility-maximizing gives economists a vocabulary for bringing these elements together.

Not even economists believe that people walk around mumbling about their marginal utilities before they walk into a shopping mall, accept a job, or make a deposit in a savings account. However, economists do believe that individuals seek their own satisfaction or utility and that people often decide to try a little less of one thing and a little more of another. If these assumptions are accepted, then the idea of utility-maximizing households facing budget constraints becomes highly plausible.

Behavioral Economics: An Alternative Viewpoint

As we know, people sometimes make decisions that seem "irrational" and not in their own best interest. People's decisions can seem inconsistent from one day to the next and they even deliberately ignore ways to save money or time. The traditional economic models assume rationality, which means that people take all available information and make consistent and informed decisions that are in their best interest. (In fact, economics professors often delight in pointing out so-called "irrational behavior" each semester to their new students, and present economics as a way to become more rational.)

But a new group of economists, known as behavioral economists, argue that the traditional method leaves out something important: people's state of mind. For example, one can think differently about money if one is feeling revenge, optimism, or loss. These are not necessarily irrational states of mind, but part of a range of emotions that can affect anyone on a given day. And what's more, actions under these conditions are indeed predictable, if the underlying environment is better understood. So, **behavioral economics** seeks to enrich the understanding of decision-making by integrating the insights of psychology into economics. It does this by investigating how given dollar amounts can mean different things to individuals depending on the situation. This can lead to decisions that appear outwardly inconsistent, or irrational, to the outside observer.

The way the mind works, according to this view, may seem inconsistent to traditional economists but is actually far more complex than an unemotional cost-benefit adding machine. For example, a traditional economist would say that if you lost a $10 bill today, and also got an extra $10 in your paycheck, you should feel perfectly neutral. After all, –$10 + $10 = $0. You are the same financially as you were before. However, behavioral economists have done research that shows many people will feel some negative emotion—anger, frustration, and so forth—after those two things happen. We tend to focus more on the loss than the gain. This is known as loss aversion, where a $1 loss pains us 2.25 times more than a $1 gain helps us, according to the economists Daniel Kahneman and Amos Tversky in a famous 1979 article in the journal *Econometrica*. This insight has implications for investing, as people tend to "overplay" the stock market by reacting more to losses than to gains. Indeed, this behavior looks irrational to traditional economists, but is consistent once we understand better how the mind works, these economists argue.

Traditional economists also assume human beings have complete self-control. But, for instance, people will buy cigarettes by the pack instead of the carton even though the carton saves them money, to keep usage down. They purchase locks for their refrigerators and overpay on taxes to force themselves to save. In other words, we protect ourselves from our worst temptations but pay a price to do so. One way behavioral economists are responding to this is by setting up ways for people to keep themselves free of these temptations. This includes what are called "nudges" toward more rational behavior rather than mandatory regulations from government. For example, up to 20 percent of new employees do not enroll in retirement savings plans immediately, because of procrastination or feeling overwhelmed by the different choices. Some companies are now moving to a new system, where employees are automatically enrolled unless they "opt out." Almost no-one opts out in this program and employees begin saving at the early years, which are most critical for retirement.

Another area that seems illogical is the idea of mental accounting, or putting dollars in different mental categories where they take different values. Economists typically consider dollars to be **fungible**, or having equal value to the individual, regardless of the situation.

You might, for instance, think of the $25 you found in the street differently from the $25 you earned from three hours working in a fast food restaurant. The street money might well be treated as "mad money" with little rational regard to getting the best value. This is in one sense strange, since it is still equivalent to three hours of hard work in the restaurant. Yet the "easy come-easy go" mentality replaces the rational economizer because of the situation, or context, in which the money was attained.

In another example of mental accounting that seems inconsistent to a traditional economist, a person could carry a credit card debt of $1,000 that has a 15% yearly interest cost, and simultaneously have a $2,000 savings account that pays only 2% per year. That means she pays $150 a year to the credit card company, while collecting only $40 annually in bank interest, so she loses $130 a year. That doesn't seem wise.

The "rational" decision would be to pay off the debt, since a $1,000 savings account with $0 in debt is the equivalent net worth, and she would now net $20 per year. But curiously, it is not uncommon for people to ignore this advice, since they will treat a loss to their savings account as higher than the benefit of paying off their credit card. The dollars are not being treated as fungible so it looks irrational to traditional economists.

Which view is right, the behavioral economists' or the traditional view? Both have their advantages, but behavioral economists have at least shed a light on trying to describe and explain behavior that has historically been dismissed as irrational. If most of us are engaged in some "irrational behavior," perhaps there are deeper underlying reasons for this behavior in the first place.

Bring it Home

"Eeny, Meeny, Miney, Moe"—Making Choices

In what category did consumers worldwide increase their spending during the recession? Higher education. According to the United Nations Educational, Scientific, and Cultural Organization (UNESCO), enrollment in colleges and universities rose one-third in China and almost two-thirds in Saudi Arabia, nearly doubled in Pakistan, tripled in Uganda, and surged by three million—18 percent—in the United States. Why were consumers willing to spend on education during lean times? Both individuals and countries view higher education as the way to prosperity. Many feel that increased earnings are a significant benefit of attending college.

Bureau of Labor Statistics data from May 2012 supports this view, as shown in **Figure 6.10**. They show a positive correlation between earnings and education. The data also indicate that unemployment rates fall with higher levels of education and training.

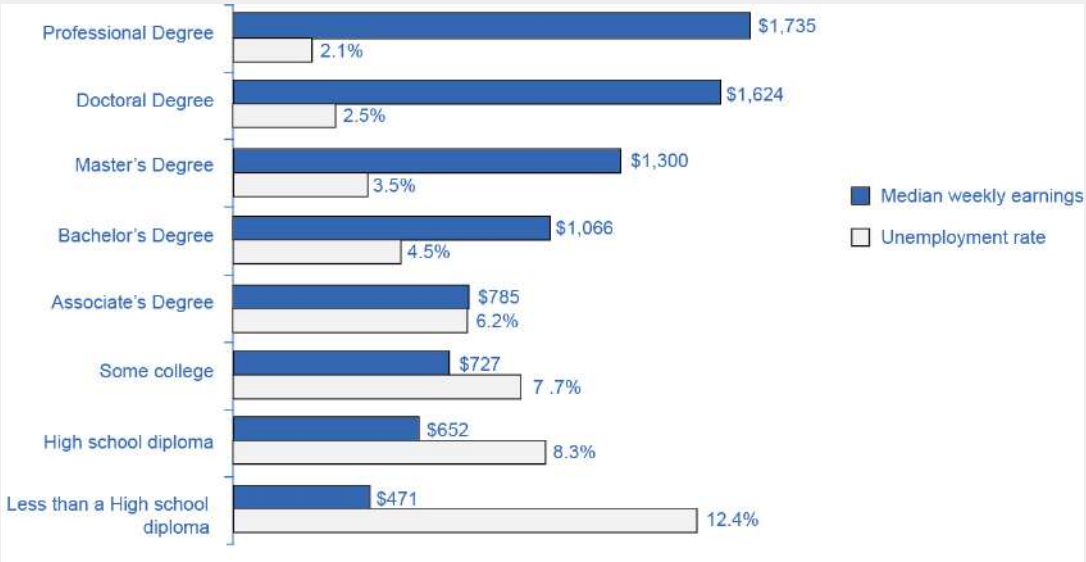


Figure 6.10 The Impact of Education on Earnings and Unemployment Rates, 2012 Those with the highest degrees in 2012 had substantially lower unemployment rates whereas those with the least formal education suffered from the highest unemployment rates. The national median average weekly income was $815, and the nation unemployment average in 2012 was 6.8%. (Source: Bureau of Labor Statistics, May 22, 2013)

KEY TERMS

backward-bending supply curve for labor the situation when high-wage people can earn so much that they respond to a still-higher wage by working fewer hours

behavioral economics a branch of economics that seeks to enrich the understanding of decision-making by integrating the insights of psychology and by investigating how given dollar amounts can mean different things to individuals depending on the situation.

budget constraint (or budget line) shows the possible combinations of two goods that are affordable given a consumer's limited income

consumer equilibrium point on the budget line where the consumer gets the most satisfaction; this occurs when the ratio of the prices of goods is equal to the ratio of the marginal utilities.

diminishing marginal utility the common pattern that each marginal unit of a good consumed provides less of an addition to utility than the previous unit

fungible the idea that units of a good, such as dollars, ounces of gold, or barrels of oil are capable of mutual substitution with each other and carry equal value to the individual.

income effect a higher price means that, in effect, the buying power of income has been reduced, even though actual income has not changed; always happens simultaneously with a substitution effect

marginal utility the additional utility provided by one additional unit of consumption

marginal utility per dollar the additional satisfaction gained from purchasing a good given the price of the product; MU/Price

substitution effect when a price changes, consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price; always happens simultaneously with an income effect

total utility satisfaction derived from consumer choices

KEY CONCEPTS AND SUMMARY

6.1 Consumption Choices

Economic analysis of household behavior is based on the assumption that people seek the highest level of utility or satisfaction. Individuals are the only judge of their own utility. In general, greater consumption of a good brings higher total utility. However, the additional utility people receive from each unit of greater consumption tends to decline in a pattern of diminishing marginal utility.

We can find the utility-maximizing choice on a consumption budget constraint in several ways. You can add up total utility of each choice on the budget line and choose the highest total. You can select a starting point at random and compare the marginal utility gains and losses of moving to neighboring points—and thus eventually seek out the preferred choice. Alternatively, you can compare the ratio of the marginal utility to price of good 1 with the marginal utility to price of good 2 and apply the rule that at the optimal choice, the two ratios should be equal:

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

6.2 How Changes in Income and Prices Affect Consumption Choices

The budget constraint framework suggest that when income or price changes, a range of responses are possible. When income rises, households will demand a higher quantity of normal goods, but a lower quantity of inferior goods. When

the price of a good rises, households will typically demand less of that good—but whether they will demand a much lower quantity or only a slightly lower quantity will depend on personal preferences. Also, a higher price for one good can lead to more or less demand of the other good.

6.3 Labor-Leisure Choices

When making a choice along the labor-leisure budget constraint, a household will choose the combination of labor, leisure, and income that provides the most utility. The result of a change in wage levels can be higher work hours, the same work hours, or lower work hours.

6.4 Intertemporal Choices in Financial Capital Markets

When making a choice along the intertemporal budget constraint, a household will choose the combination of present consumption, savings, and future consumption that provides the most utility. The result of a higher rate of return (or higher interest rates) can be a higher quantity of saving, the same quantity of saving, or a lower quantity of saving, depending on preferences about present and future consumption. Behavioral economics is a branch of economics that seeks to understand and explain the "human" factors that drive what traditional economists see as people's irrational spending decisions.

SELF-CHECK QUESTIONS

1. Jeremy is deeply in love with Jasmine. Jasmine lives where cell phone coverage is poor, so he can either call her on the land-line phone for five cents per minute or he can drive to see her, at a round-trip cost of $2 in gasoline money. He has a total of $10 per week to spend on staying in touch. To make his preferred choice, Jeremy uses a handy utilimometer that measures his total utility from personal visits and from phone minutes. Using the values in **Table 6.11**, figure out the points on Jeremy's consumption choice budget constraint (it may be helpful to do a sketch) and identify his utility-maximizing point.

Round Trips	Total Utility	Phone Minutes	Total Utility
0	0	0	0
1	80	20	200
2	150	40	380
3	210	60	540
4	260	80	680
5	300	100	800
6	330	120	900
7	200	140	980
8	180	160	1040
9	160	180	1080
10	140	200	1100

Table 6.11

2. Take Jeremy's total utility information in **Exercise 6.1**, and use the marginal utility approach to confirm the choice of phone minutes and round trips that maximize Jeremy's utility.

3. Explain all the reasons why a decrease in a product's price would lead to an increase in purchases.

4. As a college student you work at a part-time job, but your parents also send you a monthly "allowance." Suppose one month your parents forgot to send the check. Show graphically how your budget constraint is affected. Assuming you only buy normal goods, what would happen to your purchases of goods?

5. Siddhartha has 50 hours per week to devote to work or leisure. He has been working for $8 per hour. Based on the information in **Table 6.12**, calculate his utility-maximizing choice of labor and leisure time.

Leisure Hours	Total Utility from Leisure	Work Hours	Income	Total Utility from Income
0	0	0	0	0
10	200	10	80	500
20	350	20	160	800
30	450	30	240	1,040
40	500	40	320	1,240
50	530	50	400	1,400

Table 6.12

6. In Siddhartha's problem, calculate marginal utility for income and for leisure. Now, start off at the choice with 50 hours of leisure and zero income, and a wage of $8 per hour, and explain, in terms of marginal utility how Siddhartha could reason his way to the optimal choice, using marginal thinking only.

7. How would an increase in expected income over one's lifetime affect one's intertemporal budget constraint? How would it affect one's consumption/saving decision?

8. How would a decrease in expected interest rates over one's working life affect one's intertemporal budget constraint? How would it affect one's consumption/saving decision?

REVIEW QUESTIONS

9. Who determines how much utility an individual will receive from consuming a good?

10. Would you expect total utility to rise or fall with additional consumption of a good? Why?

11. Would you expect marginal utility to rise or fall with additional consumption of a good? Why?

12. Is it possible for total utility to increase while marginal utility diminishes? Explain.

13. If people do not have a complete mental picture of total utility for every level of consumption, how can they find their utility-maximizing consumption choice?

14. What is the rule relating the ratio of marginal utility to prices of two goods at the optimal choice? Explain why, if this rule does not hold, the choice cannot be utility-maximizing.

15. As a general rule, is it safe to assume that a change in the price of a good will always have its most significant impact on the quantity demanded of that good, rather than on the quantity demanded of other goods? Explain.

16. Why does a change in income cause a parallel shift in the budget constraint?

17. How will a utility-maximizer find the choice of leisure and income that provides the greatest utility?

18. As a general rule, is it safe to assume that a higher wage will encourage significantly more hours worked for all individuals? Explain.

19. According to the model of intertemporal choice, what are the major factors which determine how much saving an individual will do? What factors might a behavioral economist use to explain savings decisions?

20. As a general rule, is it safe to assume that a lower interest rate will encourage significantly lower financial savings for all individuals? Explain.

CRITICAL THINKING QUESTIONS

21. Think back to a purchase that you made recently. How would you describe your thinking before you made that purchase?

22. The rules of politics are not always the same as the rules of economics. In discussions of setting budgets for government agencies, there is a strategy called "closing the Washington Monument." When an agency faces the unwelcome prospect of a budget cut, it may decide to close a high-visibility attraction enjoyed by many people (like the Washington Monument). Explain in terms of diminishing marginal utility why the Washington Monument strategy is so misleading. *Hint*: If you are really trying to make the best of a budget cut, should you cut the items in your budget with the highest marginal utility or the lowest marginal utility? Does the Washington Monument strategy cut the items with the highest marginal utility or the lowest marginal utility?

23. Income effects depend on the income elasticity of demand for each good that you buy. If one of the goods you buy has a negative income elasticity, that is, it is an inferior good, what must be true of the income elasticity of the other good you buy?

24. In the labor-leisure choice model, what is the price of leisure?

25. Think about the backward-bending part of the labor supply curve. Why would someone work less as a result of a higher wage rate?

26. What would be the substitution effect and the income effect of a wage increase?

27. Visit the BLS website and determine if education level, race/ethnicity, or gender appear to impact labor versus leisure choices.

28. What do you think accounts for the wide range of savings rates in different countries?

29. What assumptions does the model of intertemporal choice make that are not likely true in the real world and would make the model harder to use in practice?

PROBLEMS

30. Praxilla, who lived in ancient Greece, derives utility from reading poems and from eating cucumbers. Praxilla gets 30 units of marginal utility from her first poem, 27 units of marginal utility from her second poem, 24 units of marginal utility from her third poem, and so on, with marginal utility declining by three units for each additional poem. Praxilla gets six units of marginal utility for each of her first three cucumbers consumed, five units of marginal utility for each of her next three cucumbers consumed, four units of marginal utility for each of the following three cucumbers consumed, and so on, with marginal utility declining by one for every three cucumbers consumed. A poem costs three bronze coins but a cucumber costs only one bronze coin. Praxilla has 18 bronze coins. Sketch Praxilla's budget set between poems and cucumbers, placing poems on the vertical axis and cucumbers on the horizontal axis. Start off with the choice of zero poems and 18 cucumbers, and calculate the changes in marginal utility of moving along the budget line to the next choice of one poem and 15 cucumbers. Using this step-by-step process based on marginal utility, create a table and identify Praxilla's utility-maximizing choice. Compare the marginal utility of the two goods and the relative prices at the optimal choice to see if the expected relationship holds. *Hint*: Label the table columns: 1) Choice, 2) Marginal Gain from More Poems, 3) Marginal Loss from Fewer Cucumbers, 4) Overall Gain or Loss, 5) Is the previous choice optimal? Label the table rows: 1) 0 Poems and 18 Cucumbers, 2) 1 Poem and 15 Cucumbers, 3) 2 Poems and 12 Cucumbers, 4) 3 Poems and 9 Cucumbers, 5) 4 Poems and 6 Cucumbers, 6) 5 Poems and 3 Cucumbers, 7) 6 Poems and 0 Cucumbers.

31. If a 10% decrease in the price of one product that you buy causes an 8% increase in quantity demanded of that product, will another 10% decrease in the price cause another 8% increase (no more and no less) in quantity demanded?

7 | Production, Costs and Industry Structure

Figure 7.1 Amazon is an American international electronic commerce company that sells books, among many other things, shipping them directly to the consumer. Until recently there were no brick and mortar Amazon stores. (Credit: modification of work by William Christiansen/Flickr Creative Commons)

Bring it Home

Amazon

In less than two decades, Amazon.com has transformed the way consumers sell, buy, and even read. Prior to Amazon, independent bookstores with limited inventories in small retail locations primarily sold books. There were exceptions, of course. Borders and Barnes & Noble offered larger stores in urban areas. In the last decade, however, independent bookstores have mostly disappeared, Borders has gone out of business, and Barnes & Noble is struggling. Online delivery and purchase of books has overtaken the more traditional business models. How has Amazon changed the book selling industry? How has it managed to crush its competition?

A major reason for the giant retailer's success is its production model and cost structure, which has enabled Amazon to undercut the competitors' prices even when factoring in the cost of shipping. Read on to see how firms great (like Amazon) and small (like your corner deli) determine what to sell, at what output, and price.

Introduction to Production, Costs, and Industry Structure

In this chapter, you will learn about:

- Explicit and Implicit Costs, and Accounting and Economic Profit
- Production in the Short Run

- Costs in the Short Run
- Production in the Long Run
- Costs in the Long Run

This chapter is the first of four chapters that explores the *theory of the firm*. This theory explains how firms behave. What does that mean? Let's define what we mean by the firm. A **firm** (or producer or business) combines inputs of labor, capital, land, and raw or finished component materials to produce outputs. If the firm is successful, the outputs are more valuable than the inputs. This activity of **production** goes beyond manufacturing (i.e., making things). It includes any process or service that creates value, including transportation, distribution, wholesale and retail sales.

Production involves a number of important decisions that define a firm's behavior. These decisions include, but are not limited to:

- What product or products should the firm produce?
- How should the firm produce the products (i.e., what production process should the firm use)?
- How much output should the firm produce?
- What price should the firm charge for its products?
- How much labor should the firm employ?

The answers to these questions depend on the production and cost conditions facing each firm. That is the subject of this chapter. The answers also depend on the market structure for the product(s) in question. Market structure is a multidimensional concept that involves how competitive the industry is. We define it by questions such as these:

- How much market power does each firm in the industry possess?
- How similar is each firm's product to the products of other firms in the industry?
- How difficult is it for new firms to enter the industry?
- Do firms compete on the basis of price, advertising, or other product differences?

Figure 7.2 illustrates the range of different market structures, which we will explore in **Perfect Competition**, **Monopoly**, and **Monopolistic Competition and Oligopoly**.

Figure 7.2 The Spectrum of Competition Firms face different competitive situations. At one extreme—perfect competition—many firms are all trying to sell identical products. At the other extreme—monopoly—only one firm is selling the product, and this firm faces no competition. Monopolistic competition and oligopoly fall between the extremes of perfect competition and monopoly. Monopolistic competition is a situation with many firms selling similar, but not identical products. Oligopoly is a situation with few firms that sell identical or similar products.

Let's examine how firms determine their costs and desired profit levels. Then we will discuss the origins of cost, both in the short and long run. Private enterprise, which can be private individual or group business ownership, characterizes the U.S. economy. In the U.S. system, we have the option to organize private businesses as sole proprietorships (one owner), partners (more than one owner), and corporations (legal entitles separate from the owners.

When people think of businesses, often corporate giants like Wal-Mart, Microsoft, or General Motors come to mind. However, firms come in all sizes, as **Table 7.1** shows. The vast majority of American firms have fewer than 20 employees. As of 2010, the U.S. Census Bureau counted 5.7 million firms with employees in the U.S. economy.

Slightly less than half of all the workers in private firms are at the 17,000 large firms, meaning they employ more than 500 workers. Another 35% of workers in the U.S. economy are at firms with fewer than 100 workers. These small-scale businesses include everything from dentists and lawyers to businesses that mow lawns or clean houses. Table 7.1 does not include a separate category for the millions of small "non-employer" businesses where a single owner or a few partners are not officially paid wages or a salary, but simply receive whatever they can earn.

Number of Employees	Firms (% of total firms)	Number of Paid Employees (% of total employment)
Total	**5,734,538**	**112.0 million**
0–9	4,543,315 (79.2%)	12.3 million (11.0%)
10–19	617,089 (10.8%)	8.3 million (7.4%)
20–99	475,125 (8.3%)	18.6 million (16.6%)
100–499	81,773 (1.4%)	15.9 million (14.2%)
500 or more	17,236 (0.30%)	50.9 million (49.8%)

Table 7.1 Range in Size of U.S. Firms (Source: U.S. Census, 2010 www.census.gov)

7.1 | Explicit and Implicit Costs, and Accounting and Economic Profit

By the end of this section, you will be able to:

- Explain the difference between explicit costs and implicit costs
- Understand the relationship between cost and revenue

Each business, regardless of size or complexity, tries to earn a profit:

$$\text{Profit} = \text{Total Revenue} - \text{Total Cost}$$

Total **revenue** is the income the firm generates from selling its products. We calculate it by multiplying the price of the product times the quantity of output sold:

$$\text{Total Revenue} = \text{Price} \times \text{Quantity}$$

We will see in the following chapters that revenue is a function of the demand for the firm's products.

Total cost is what the firm pays for producing and selling its products. Recall that production involves the firm converting inputs to outputs. Each of those inputs has a cost to the firm. The sum of all those costs is total cost. We will learn in this chapter that short run costs are different from long run costs.

We can distinguish between two types of cost: explicit and implicit. **Explicit costs** are out-of-pocket costs, that is, actual payments. Wages that a firm pays its employees or rent that a firm pays for its office are explicit costs. **Implicit costs** are more subtle, but just as important. They represent the opportunity cost of using resources that the firm already owns. Often for small businesses, they are resources that the owners contribute. For example, working in the business while not earning a formal salary, or using the ground floor of a home as a retail store are both implicit costs. Implicit costs also include the depreciation of goods, materials, and equipment that are necessary for a company to operate. (See the Work It Out feature for an extended example.)

These two definitions of cost are important for distinguishing between two conceptions of profit, accounting profit, and economic profit. **Accounting profit** is a cash concept. It means total revenue minus explicit costs—the difference between dollars brought in and dollars paid out. **Economic profit** is total revenue minus total cost, including both explicit and implicit costs. The difference is important because even though a business pays income taxes based on

its accounting profit, whether or not it is economically successful depends on its economic profit.

Work It Out

Calculating Implicit Costs

Consider the following example. Fred currently works for a corporate law firm. He is considering opening his own legal practice, where he expects to earn $200,000 per year once he establishes himself. To run his own firm, he would need an office and a law clerk. He has found the perfect office, which rents for $50,000 per year. He could hire a law clerk for $35,000 per year. If these figures are accurate, would Fred's legal practice be profitable?

Step 1. First you have to calculate the costs. You can take what you know about explicit costs and total them:

Office rental :	$50,000
Law clerk's salary :	+$35,000
Total explicit costs :	$85,000

Step 2. Subtracting the explicit costs from the revenue gives you the accounting profit.

Revenues :	$200,000
Explicit costs :	–$85,000
Accounting profit :	$115,000

However, these calculations consider only the explicit costs. To open his own practice, Fred would have to quit his current job, where he is earning an annual salary of $125,000. This would be an implicit cost of opening his own firm.

Step 3. You need to subtract both the explicit and implicit costs to determine the true economic profit:

$$\begin{aligned} \text{Economic profit} &= \text{total revenues} - \text{explicit costs} - \text{implicit costs} \\ &= \$200{,}000 - \$85{,}000 - \$125{,}000 \\ &= -\$10{,}000 \text{ per year} \end{aligned}$$

Fred would be losing $10,000 per year. That does not mean he would not want to open his own business, but it does mean he would be earning $10,000 less than if he worked for the corporate firm.

Implicit costs can include other things as well. Maybe Fred values his leisure time, and starting his own firm would require him to put in more hours than at the corporate firm. In this case, the lost leisure would also be an implicit cost that would subtract from economic profits.

Now that we have an idea about the different types of costs, let's look at cost structures. A firm's cost structure in the long run may be different from that in the short run. We turn to that distinction in the next few sections.

7.2 | Production in the Short Run

By the end of this section, you will be able to:

- Understand the concept of a production function
- Differentiate between the different types of inputs or factors in a production function
- Differentiate between fixed and variable inputs
- Differentiate between production in the short run and in the long run
- Differentiate between total and marginal product
- Understand the concept of diminishing marginal productivity

In this chapter, we want to explore the relationship between the quantity of output a firm produces, and the cost of

producing that output. We mentioned that the cost of the product depends on how many inputs are required to produce the product and what those inputs cost. We can answer the former question by looking at the firm's production function.

Figure 7.3 The production process for pizza includes inputs such as ingredients, the efforts of the pizza maker, and tools and materials for cooking and serving. (Credit: Haldean Brown/Flickr Creative Commons)

Production is the process (or processes) a firm uses to transform inputs (e.g. labor, capital, raw materials) into outputs, i.e. the goods or services the firm wishes to sell. Consider pizza making. The pizzaiolo (pizza maker) takes flour, water, and yeast to make dough. Similarly, the pizzaiolo may take tomatoes, spices, and water to make pizza sauce. The cook rolls out the dough, brushes on the pizza sauce, and adds cheese and other toppings. The pizzaiolo uses a peel—the shovel-like wooden tool-- to put the pizza into the oven to cook. Once baked, the pizza goes into a box (if it's for takeout) and the customer pays for the good. What are the inputs (or factors of production) in the production process for this pizza?

Economists divide factors of production into several categories:

- **Natural Resources (Land and Raw Materials)** - The ingredients for the pizza are raw materials. These include the flour, yeast, and water for the dough, the tomatoes, herbs, and water for the sauce, the cheese, and the toppings. If the pizza place uses a wood-burning oven, we would include the wood as a raw material. If the establishment heats the oven with natural gas, we would count this as a raw material. Don't forget electricity for lights. If, instead of pizza, we were looking at an agricultural product, like wheat, we would include the land the farmer used for crops here.
- **Labor** – When we talk about production, labor means human effort, both physical and mental. The pizzaiolo was the primary example of labor here. He or she needs to be strong enough to roll out the dough and to insert and retrieve the pizza from the oven, but he or she also needs to know **how** to make the pizza, how long it cooks in the oven and a myriad of other aspects of pizza-making. The business may also have one or more people to work the counter, take orders, and receive payment.
- **Capital** – When economists uses the term capital, they do not mean financial capital (money); rather, they mean physical capital, the machines, equipment, and buildings that one uses to produce the product. In the case of pizza, the capital includes the peel, the oven, the building, and any other necessary equipment (for example, tables and chairs).
- **Technology** – Technology refers to the process or processes for producing the product. How does the pizzaiolo combine ingredients to make pizza? How hot should the oven be? How long should the pizza cook? What is the best oven to use? Gas or wood burning? Should the restaurant make its own dough, sauce, cheese, toppings, or should it buy them?
- **Entrepreneurship** – Production involves many decisions and much knowledge, even for something as simple as pizza. Who makes those decisions? Ultimately, it is the entrepreneur, the person who creates the business,

whose idea it is to combine the inputs to produce the outputs.

The cost of producing pizza (or any output) depends on the amount of labor capital, raw materials, and other inputs required and the price of each input to the entrepreneur. Let's explore these ideas in more detail.

We can summarize the ideas so far in terms of a **production function**, a mathematical expression or equation that explains the engineering relationship between inputs and outputs:

$$Q = f[NR, L, K, t, E]$$

The production function gives the answer to the question, how much output can the firm produce given different amounts of inputs? Production functions are specific to the product. Different products have different production functions. The amount of labor a farmer uses to produce a bushel of wheat is likely different than that required to produce an automobile. Firms in the same industry may have somewhat different production functions, since each firm may produce a little differently. One pizza restaurant may make its own dough and sauce, while another may buy those pre-made. A sit-down pizza restaurant probably uses more labor (to handle table service) than a purely take-out restaurant.

We can describe inputs as either **fixed** or **variable**.

Fixed inputs are those that can't easily be increased or decreased in a short period of time. In the pizza example, the building is a fixed input. Once the entrepreneur signs the lease, he or she is stuck in the building until the lease expires. Fixed inputs define the firm's maximum output capacity. This is analogous to the potential real GDP shown by society's production possibilities curve, i.e. the maximum quantities of outputs a society can produce at a given time with its available resources.

Variable inputs are those that can easily be increased or decreased in a short period of time. The pizzaiolo can order more ingredients with a phone call, so ingredients would be variable inputs. The owner could hire a new person to work the counter pretty quickly as well.

Economists often use a short-hand form for the production function:

$$Q = f[L, K],$$

where L represents all the variable inputs, and K represents all the fixed inputs.

Economists differentiate between short and long run production.

The **short run** is the period of time during which at least some factors of production are fixed. During the period of the pizza restaurant lease, the pizza restaurant is operating in the short run, because it is limited to using the current building—the owner can't choose a larger or smaller building.

The **long run** is the period of time during which all factors are variable. Once the lease expires for the pizza restaurant, the shop owner can move to a larger or smaller place.

Let's explore production in the short run using a specific example: tree cutting (for lumber) with a two-person crosscut saw.

Figure 7.4 Production in the short run may be explored through the example of lumberjacks using a two-person saw. (Credit: Wknight94/Wikimedia Commons)

Since by definition capital is fixed in the short run, our production function becomes

$$Q = f[L, \bar{K}] \text{ or } Q = f[L]$$

This equation simply indicates that since capital is fixed, the amount of output (e.g. trees cut down per day) depends only on the amount of labor employed (e.g. number of lumberjacks working). We can express this production function numerically as **Table 7.2** below shows.

# Lumberjacks	1	2	3	4	5
# Trees (TP)	4	10	12	13	13
MP	4	6	2	1	0

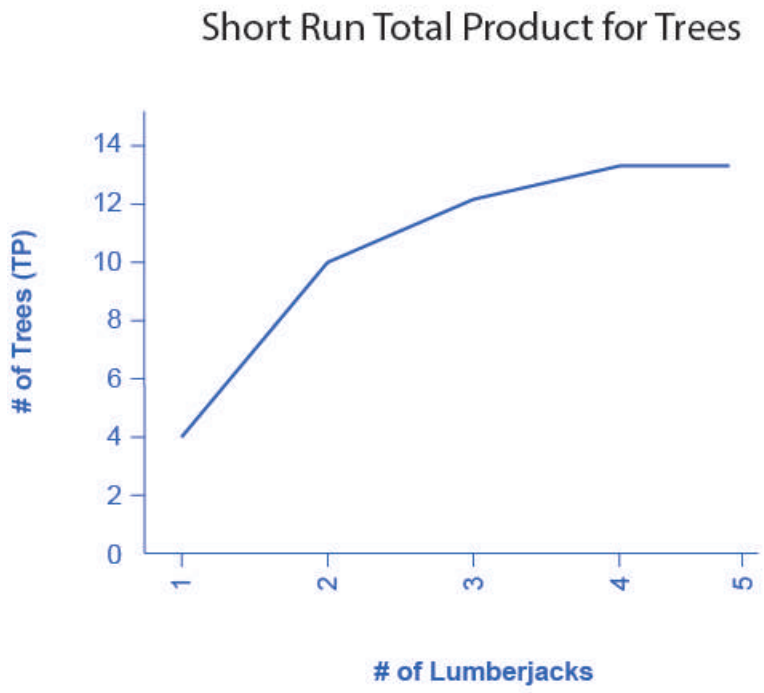
Table 7.2 Short Run Production Function for Trees

Note that we have introduced some new language. We also call Output (Q) Total Product (TP), which means the amount of output produced with a given amount of labor and a fixed amount of capital. In this example, one lumberjack using a two-person saw can cut down four trees in an hour. Two lumberjacks using a two-person saw can cut down ten trees in an hour.

We should also introduce a critical concept: **marginal product**. Marginal product is the additional output of one more worker. Mathematically, Marginal Product is the change in total product divided by the change in labor: $MP = \Delta TP / \Delta L$. In the table above, since 0 workers produce 0 trees, the marginal product of the first worker is four trees per day, but the marginal product of the second worker is six trees per day. Why might that be the case? It's because of the nature of the capital the workers are using. A two-person saw works much better with two persons than with one. Suppose we add a third lumberjack to the story. What will that person's marginal product be? What will that person contribute to the team? Perhaps he or she can oil the saw's teeth to keep it sawing smoothly or he or she could

bring water to the two people sawing. What you see in the table is a critically important conclusion about production in the short run: It may be that as we add workers, the marginal product increases at first, but sooner or later additional workers will have decreasing marginal product. In fact, there may eventually be no effect or a negative effect on output. This is called the **Law of Diminishing Marginal Product** and it's a characteristic of production in the short run. Diminishing marginal productivity is very similar to the concept of diminishing marginal utility that we learned about in the chapter on consumer choice. Both concepts are examples of the more general concept of diminishing marginal returns. Why does diminishing marginal productivity occur? It's because of fixed capital. We will see this more clearly when we discuss production in the long run.

We can show these concepts graphically as **Figure 7.5** and **Figure 7.6** illustrate. **Figure 7.5** graphically shows the data from **Table 7.2**. **Figure 7.6** shows the more general cases of total product and marginal product curves.


Short Run Total Product for Trees
# of Trees (TP)
0
2
4
6
8
10
12
14
1
2
3
4
5
# of Lumberjacks

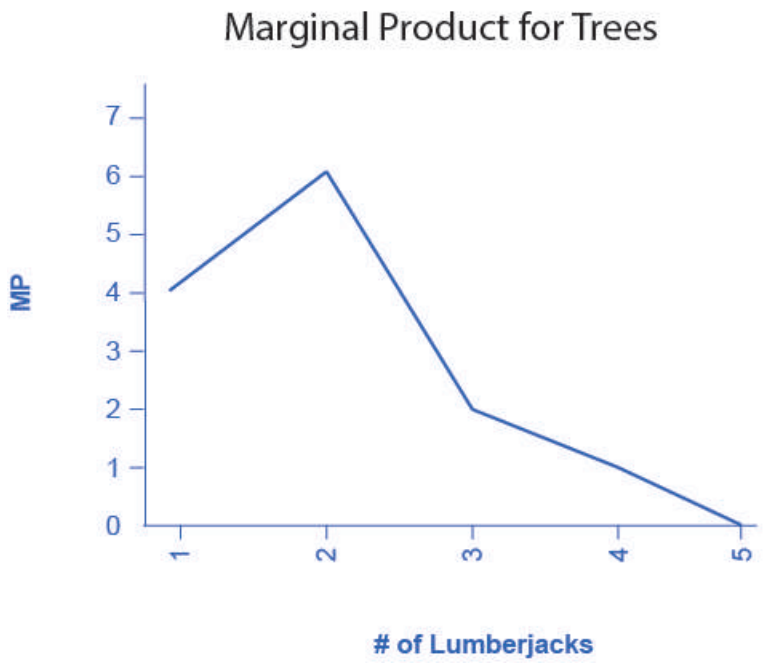

Marginal Product for Trees
MP
0
1
2
3
4
5
6
7
1
2
3
4
5
# of Lumberjacks

Figure 7.5

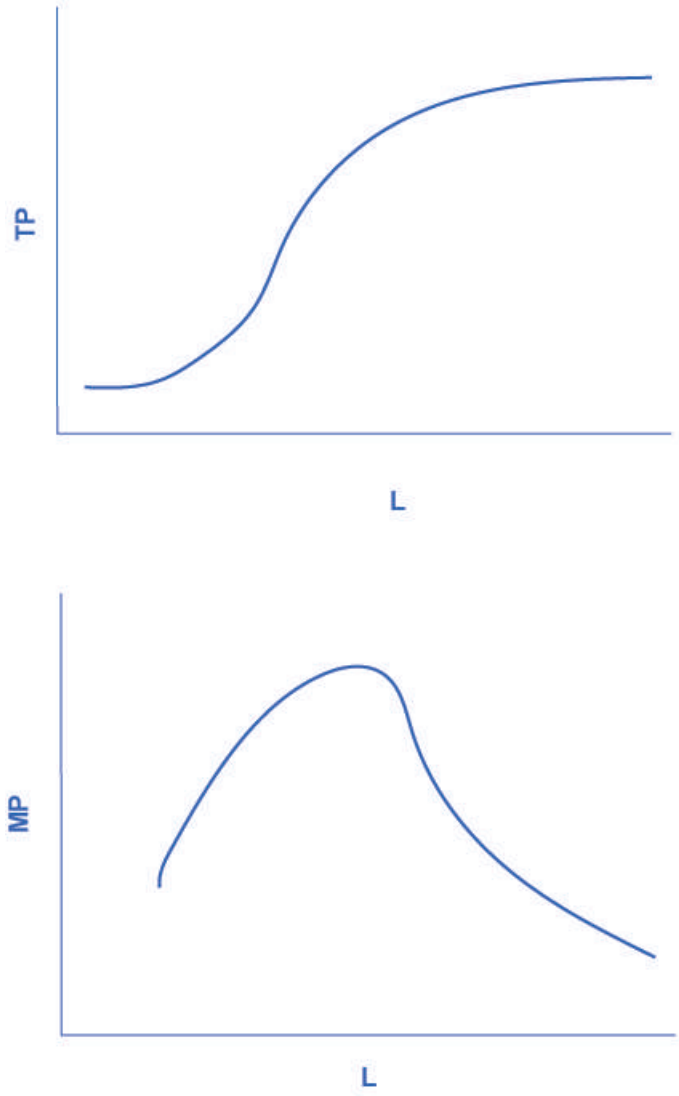


Figure 7.6

7.3 | Costs in the Short Run

By the end of this section, you will be able to:

- Understand the relationship between production and costs
- Understand that every factor of production has a corresponding factor price
- Analyze short-run costs in terms of total cost, fixed cost, variable cost, marginal cost, and average cost
- Calculate average profit
- Evaluate patterns of costs to determine potential profit

We've explained that a firm's total costs depend on the quantities of inputs the firm uses to produce its output and the cost of those inputs to the firm. The firm's production function tells us how much output the firm will produce with given amounts of inputs. However, if we think about that backwards, it tells us how many inputs the firm needs to produce a given quantity of output, which is the first thing we need to determine total cost. Let's move to the second

factor we need to determine.

For every factor of production (or input), there is an associated factor payment. **Factor payments** are what the firm pays for the use of the factors of production. From the firm's perspective, factor payments are costs. From the owner of each factor's perspective, factor payments are income. Factor payments include:

- **Raw materials prices** for raw materials
- **Rent** for land or buildings
- **Wages and salaries** for labor
- **Interest and dividends** for the use of financial capital (loans and equity investments)
- **Profit** for entrepreneurship. Profit is the residual, what's left over from revenues after the firm pays all the other costs. While it may seem odd to treat profit as a "cost", it is what entrepreneurs earn for taking the risk of starting a business. You can see this correspondence between factors of production and factor payments in the inside loop of the circular flow diagram in **Figure 1.6**.

We now have all the information necessary to determine a firm's costs.

A cost function is a mathematical expression or equation that shows the cost of producing different levels of output.

Q	1	2	3	4
Cost	$32.50	$44	$52	$90

Table 7.3 Cost Function for Producing Widgets

What we observe is that the cost increases as the firm produces higher quantities of output. This is pretty intuitive, since producing more output requires greater quantities of inputs, which cost more dollars to acquire.

What is the origin of these cost figures? They come from the production function and the factor payments. The discussion of costs in the short run above, **Costs in the Short Run**, was based on the following production function, which is similar to **Table 7.3** except for "widgets" instead of trees.

Workers (L)	1	2	3	3.25	4.4	5.2	6	7	8	9
Widgets (Q)	0.2	0.4	0.8	1	2	3	3.5	3.8	3.95	4

Table 7.4

We can use the information from the production function to determine production costs. What we need to know is how many workers are required to produce any quantity of output. If we flip the order of the rows, we "invert" the production function so it shows $L = f(Q)$.

Widgets (Q)	0.2	0.4	0.8	1	2	3	3.5	3.8	3.95	4
Workers (L)	1	2	3	3.25	4.4	5.2	6	7	8	9

Table 7.5

Now focus on the whole number quantities of output. We'll eliminate the fractions from the table:

Widgets (Q)				1	2	3				4

Table 7.6

Workers (L)				3.25	4.4	5.2				9

Table 7.6

Suppose widget workers receive $10 per hour. Multiplying the Workers row by $10 (and eliminating the blanks) gives us the cost of producing different levels of output.

Widgets (Q)	1.00	2.00	3.00	4.00
Workers (L)	3.25	4.4	5.2	9
× Wage Rate per hour	$10	$10	$10	$10
= Cost	$32.50	$44.00	$52.00	$90.00

Table 7.7

This is same cost function with which we began! (shown in **Table 7.3**)

Now that we have the basic idea of the cost origins and how they are related to production, let's drill down into the details.

Average and Marginal Costs

The cost of producing a firm's output depends on how much labor and physical capital the firm uses. A list of the costs involved in producing cars will look very different from the costs involved in producing computer software or haircuts or fast-food meals.

We can measure costs in a variety of ways. Each way provides its own insight into costs. Sometimes firms need to look at their cost per unit of output, not just their total cost. There are two ways to measure per unit costs. The most intuitive way is average cost. Average cost is the cost on average of producing a given quantity. We define **average cost** as total cost divided by the quantity of output produced. $AC = TC / Q$ If producing two widgets costs a total of $44, the average cost per widget is $\$44 / 2 = \$22$ per widget. The other way of measuring cost per unit is marginal cost. If average cost is the cost of the average unit of output produced, marginal cost is the cost of each individual unit produced. More formally, marginal cost is the cost of producing one more unit of output. Mathematically, **marginal cost** is the change in total cost divided by the change in output: $MC = \Delta TC / \Delta Q$. If the cost of the first widget is $32.50 and the cost of two widgets is $44, the marginal cost of the second widget is $\$44 - \$32.50 = \$11.50$. We can see the Widget Cost table redrawn below with average and marginal cost added.

Q	1	2	3	4
Total Cost	$32.50	$44.00	$52.00	$90.00
Average Cost	$32.50	$22.00	$17.33	$22.50
Marginal Cost	$32.50	$11.50	$8.00	$38.00

Table 7.8 Extended Cost Function for Producing Widgets

Note that the marginal cost of the first unit of output is always the same as total cost.

Fixed and Variable Costs

We can decompose costs into fixed and variable costs. Fixed costs are the costs of the fixed inputs (e.g. capital). Because fixed inputs do not change in the short run, fixed costs are expenditures that do not change regardless of the level of production. Whether you produce a great deal or a little, the fixed costs are the same. One example is the rent on a factory or a retail space. Once you sign the lease, the rent is the same regardless of how much you produce, at

least until the lease expires. Fixed costs can take many other forms: for example, the cost of machinery or equipment to produce the product, research and development costs to develop new products, even an expense like advertising to popularize a brand name. The amount of fixed costs varies according to the specific line of business: for instance, manufacturing computer chips requires an expensive factory, but a local moving and hauling business can get by with almost no fixed costs at all if it rents trucks by the day when needed.

Variable costs are the costs of the variable inputs (e.g. labor). The only way to increase or decrease output is by increasing or decreasing the variable inputs. Therefore, variable costs increase or decrease with output. We treat labor as a variable cost, since producing a greater quantity of a good or service typically requires more workers or more work hours. Variable costs would also include raw materials.

Total costs are the sum of fixed plus variable costs. Let's look at another example. Consider the barber shop called "The Clip Joint" in **Figure 7.7**. The data for output and costs are in **Table 7.9**. The fixed costs of operating the barber shop, including the space and equipment, are $160 per day. The variable costs are the costs of hiring barbers, which in our example is $80 per barber each day. The first two columns of the table show the quantity of haircuts the barbershop can produce as it hires additional barbers. The third column shows the fixed costs, which do not change regardless of the level of production. The fourth column shows the variable costs at each level of output. We calculate these by taking the amount of labor hired and multiplying by the wage. For example, two barbers cost: 2 × $80 = $160. Adding together the fixed costs in the third column and the variable costs in the fourth column produces the total costs in the fifth column. For example, with two barbers the total cost is: $160 + $160 = $320.

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost
1	16	$160	$80	$240
2	40	$160	$160	$320
3	60	$160	$240	$400
4	72	$160	$320	$480
5	80	$160	$400	$560
6	84	$160	$480	$640
7	82	$160	$560	$720

Table 7.9 Output and Total Costs

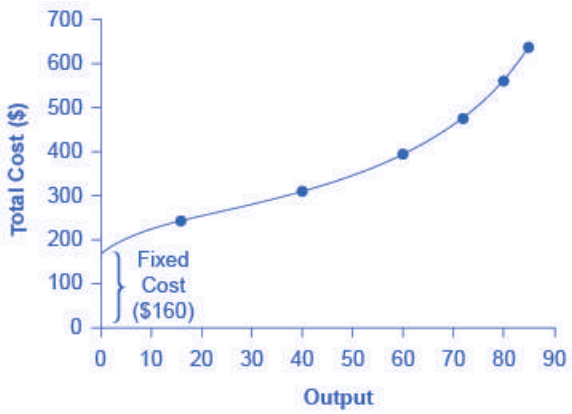


Figure 7.7 How Output Affects Total Costs At zero production, the fixed costs of $160 are still present. As production increases, variable costs are added to fixed costs, and the total cost is the sum of the two.

At zero production, the fixed costs of $160 are still present. As production increases, we add variable costs to fixed costs, and the total cost is the sum of the two. **Figure 7.7** graphically shows the relationship between the quantity of

output produced and the cost of producing that output. We always show the fixed costs as the vertical intercept of the total cost curve; that is, they are the costs incurred when output is zero so there are no variable costs.

You can see from the graph that once production starts, total costs and variable costs rise. While variable costs may initially increase at a decreasing rate, at some point they begin increasing at an increasing rate. This is caused by diminishing marginal productivity which we discussed earlier in the **Production in the Short Run** section of this chapter, which is easiest to see with an example. As the number of barbers increases from zero to one in the table, output increases from 0 to 16 for a marginal gain (or marginal product) of 16. As the number rises from one to two barbers, output increases from 16 to 40, a marginal gain of 24. From that point on, though, the marginal product diminishes as we add each additional barber. For example, as the number of barbers rises from two to three, the marginal product is only 20; and as the number rises from three to four, the marginal product is only 12.

To understand the reason behind this pattern, consider that a one-man barber shop is a very busy operation. The single barber needs to do everything: say hello to people entering, answer the phone, cut hair, sweep, and run the cash register. A second barber reduces the level of disruption from jumping back and forth between these tasks, and allows a greater division of labor and specialization. The result can be increasing marginal productivity. However, as the shop adds other barbers, the advantage of each additional barber is less, since the specialization of labor can only go so far. The addition of a sixth or seventh or eighth barber just to greet people at the door will have less impact than the second one did. This is the pattern of diminishing marginal productivity. As a result, the total costs of production will begin to rise more rapidly as output increases. At some point, you may even see negative returns as the additional barbers begin bumping elbows and getting in each other's way. In this case, the addition of still more barbers would actually cause output to decrease, as the last row of **Table 7.9** shows.

This pattern of diminishing marginal productivity is common in production. As another example, consider the problem of irrigating a crop on a farmer's field. The plot of land is the fixed factor of production, while the water that the farmer can add to the land is the key variable cost. As the farmer adds water to the land, output increases. However, adding increasingly more water brings smaller increases in output, until at some point the water floods the field and actually reduces output. Diminishing marginal productivity occurs because, with fixed inputs (land in this example), each additional unit of input (e.g. water) contributes less to overall production.

Average Total Cost, Average Variable Cost, Marginal Cost

The breakdown of total costs into fixed and variable costs can provide a basis for other insights as well. The first five columns of **Table 7.10** duplicate the previous table, but the last three columns show average total costs, average variable costs, and marginal costs. These new measures analyze costs on a per-unit (rather than a total) basis and are reflected in the curves in **Figure 7.8**.

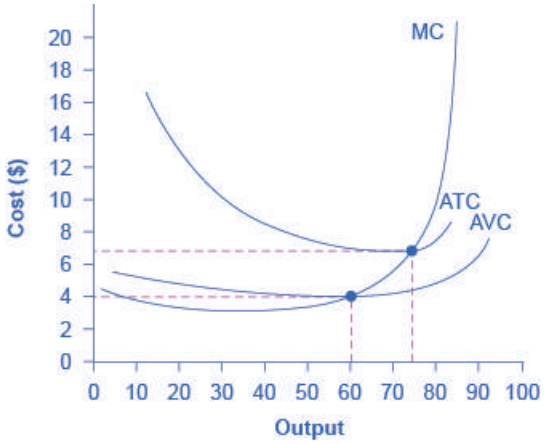


Figure 7.8 Cost Curves at the Clip Joint We can also present the information on total costs, fixed cost, and variable cost on a per-unit basis. We calculate average total cost (ATC) by dividing total cost by the total quantity produced. The average total cost curve is typically U-shaped. We calculate average variable cost (AVC) by dividing variable cost by the quantity produced. The average variable cost curve lies below the average total cost curve and is also typically U-shaped. We calculate marginal cost (MC) by taking the change in total cost between two levels of output and dividing by the change in output. The marginal cost curve is upward-sloping.

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost	Marginal Cost	Average Total Cost	Average Variable Cost
1	16	$160	$80	$240	$15.00	$15.00	$5.00
2	40	$160	$160	$320	$3.33	$8.00	$4.00
3	60	$160	$240	$400	$4.00	$6.67	$4.00
4	72	$160	$320	$480	$6.67	$6.67	$4.44
5	80	$160	$400	$560	$10.00	$7.00	$5.00
6	84	$160	$480	$640	$20.00	$7.62	$5.71

Table 7.10 Different Types of Costs

Average total cost (sometimes referred to simply as average cost) is total cost divided by the quantity of output. Since the total cost of producing 40 haircuts is $320, the average total cost for producing each of 40 haircuts is $320/40, or $8 per haircut. Average cost curves are typically U-shaped, as **Figure 7.8** shows. Average total cost starts off relatively high, because at low levels of output total costs are dominated by the fixed cost. Mathematically, the denominator is so small that average total cost is large. Average total cost then declines, as the fixed costs are spread over an increasing quantity of output. In the average cost calculation, the rise in the numerator of total costs is relatively small compared to the rise in the denominator of quantity produced. However, as output expands still further, the average cost begins to rise. At the right side of the average cost curve, total costs begin rising more rapidly as diminishing returns come into effect.

We obtain **average variable cost** when we divide variable cost by quantity of output. For example, the variable cost of producing 80 haircuts is $400, so the average variable cost is $400/80, or $5 per haircut. Note that at any level of output, the average variable cost curve will always lie below the curve for average total cost, as **Figure 7.8** shows. The reason is that average total cost includes average variable cost and average fixed cost. Thus, for Q = 80 haircuts, the average total cost is $8 per haircut, while the average variable cost is $5 per haircut. However, as output grows, fixed costs become relatively less important (since they do not rise with output), so average variable cost sneaks closer to average cost.

Average total and variable costs measure the average costs of producing some quantity of output. Marginal cost is somewhat different. **Marginal cost** is the additional cost of producing one more unit of output. It is not the cost per unit of *all* units produced, but only the next one (or next few). We calculate marginal cost by taking the change in total cost and dividing it by the change in quantity. For example, as quantity produced increases from 40 to 60 haircuts, total costs rise by 400 – 320, or 80. Thus, the marginal cost for each of those marginal 20 units will be 80/20, or $4 per haircut. The marginal cost curve is generally upward-sloping, because diminishing marginal returns implies that additional units are more costly to produce. We can see small range of increasing marginal returns in the figure as a dip in the marginal cost curve before it starts rising. There is a point at which marginal and average costs meet, as the following Clear it Up feature discusses.

Clear It Up

Where do marginal and average costs meet?

The marginal cost line intersects the average cost line exactly at the bottom of the average cost curve—which occurs at a quantity of 72 and cost of $6.60 in **Figure 7.8**. The reason why the intersection occurs at this point is built into the economic meaning of marginal and average costs. If the marginal cost of production is below the average cost for producing previous units, as it is for the points to the left of where MC crosses ATC, then producing one more additional unit will reduce average costs overall—and the ATC curve will be downward-sloping in this zone. Conversely, if the marginal cost of production for producing an additional unit is above

the average cost for producing the earlier units, as it is for points to the right of where MC crosses ATC, then producing a marginal unit will increase average costs overall—and the ATC curve must be upward-sloping in this zone. The point of transition, between where MC is pulling ATC down and where it is pulling it up, must occur at the minimum point of the ATC curve.

This idea of the marginal cost "pulling down" the average cost or "pulling up" the average cost may sound abstract, but think about it in terms of your own grades. If the score on the most recent quiz you take is lower than your average score on previous quizzes, then the marginal quiz pulls down your average. If your score on the most recent quiz is higher than the average on previous quizzes, the marginal quiz pulls up your average. In this same way, low marginal costs of production first pull down average costs and then higher marginal costs pull them up.

The numerical calculations behind average cost, average variable cost, and marginal cost will change from firm to firm. However, the general patterns of these curves, and the relationships and economic intuition behind them, will not change.

Lessons from Alternative Measures of Costs

Breaking down total costs into fixed cost, marginal cost, average total cost, and average variable cost is useful because each statistic offers its own insights for the firm.

Whatever the firm's quantity of production, total revenue must exceed total costs if it is to earn a profit. As explored in the chapter **Choice in a World of Scarcity**, fixed costs are often sunk costs that a firm cannot recoup. In thinking about what to do next, typically you should ignore sunk costs, since you have already spent this money and cannot make any changes. However, you can change variable costs, so they convey information about the firm's ability to cut costs in the present and the extent to which costs will increase if production rises.

Clear It Up

Why are total cost and average cost not on the same graph?

Total cost, fixed cost, and variable cost each reflect different aspects of the cost of production over the entire quantity of output produced. We measure these costs in dollars. In contrast, marginal cost, average cost, and average variable cost are costs per unit. In the previous example, we measured them as dollars per haircut. Thus, it would not make sense to put all of these numbers on the same graph, since we measure them in different units ($ versus $ per unit of output).

It would be as if the vertical axis measured two different things. In addition, as a practical matter, if they were on the same graph, the lines for marginal cost, average cost, and average variable cost would appear almost flat against the horizontal axis, compared to the values for total cost, fixed cost, and variable cost. Using the figures from the previous example, the total cost of producing 40 haircuts is $320. However, the average cost is $320/40, or $8. If you graphed both total and average cost on the same axes, the average cost would hardly show.

Average cost tells a firm whether it can earn profits given the current price in the market. If we divide profit by the quantity of output produced we get **average profit**, also known as the firm's *profit margin*. Expanding the equation for profit gives:

$$\begin{aligned} \text{average profit} &= \frac{\text{profit}}{\text{quantity produced}} \\ &= \frac{\text{total revenue} - \text{total cost}}{\text{quantity produced}} \\ &= \frac{\text{total revenue}}{\text{quantity produced}} - \frac{\text{total cost}}{\text{quantity produced}} \\ &= \text{average revenue} - \text{average cost} \end{aligned}$$

However, note that:

$$\text{average revenue} = \frac{\text{price} \times \text{quantity produced}}{\text{quantity produced}} = \text{price}$$

Thus:

$$\text{average profit} = \text{price} - \text{average cost}$$

This is the firm's **profit margin**. This definition implies that if the market price is above average cost, average profit, and thus total profit, will be positive. If price is below average cost, then profits will be negative.

We can compare this marginal cost of producing an additional unit with the marginal revenue gained by selling that additional unit to reveal whether the additional unit is adding to total profit—or not. Thus, marginal cost helps producers understand how increasing or decreasing production affects profits.

A Variety of Cost Patterns

The pattern of costs varies among industries and even among firms in the same industry. Some businesses have high fixed costs, but low marginal costs. Consider, for example, an internet company that provides medical advice to customers. Consumers might pay such a company directly, or perhaps hospitals or healthcare practices might subscribe on behalf of their patients. Setting up the website, collecting the information, writing the content, and buying or leasing the computer space to handle the web traffic are all fixed costs that the company must undertake before the site can work. However, when the website is up and running, it can provide a high quantity of service with relatively low variable costs, like the cost of monitoring the system and updating the information. In this case, the total cost curve might start at a high level, because of the high fixed costs, but then might appear close to flat, up to a large quantity of output, reflecting the low variable costs of operation. If the website is popular, however, a large rise in the number of visitors will overwhelm the website, and increasing output further could require a purchase of additional computer space.

For other firms, fixed costs may be relatively low. For example, consider firms that rake leaves in the fall or shovel snow off sidewalks and driveways in the winter. For fixed costs, such firms may need little more than a car to transport workers to homes of customers and some rakes and shovels. Still other firms may find that diminishing marginal returns set in quite sharply. If a manufacturing plant tried to run 24 hours a day, seven days a week, little time remains for routine equipment maintenance, and marginal costs can increase dramatically as the firm struggles to repair and replace overworked equipment.

Every firm can gain insight into its task of earning profits by dividing its total costs into fixed and variable costs, and then using these calculations as a basis for average total cost, average variable cost, and marginal cost. However, making a final decision about the profit-maximizing quantity to produce and the price to charge will require combining these perspectives on cost with an analysis of sales and revenue, which in turn requires looking at the market structure in which the firm finds itself. Before we turn to the analysis of market structure in other chapters, we will analyze the firm's cost structure from a long-run perspective.

7.4 | Production in the Long Run

By the end of this section, you will be able to:

- Understand how long run production differs from short run production.

In the long run, all factors (including capital) are variable, so our production function is $Q = f[L, K]$.

Consider a secretarial firm that does typing for hire using typists for labor and personal computers for capital. To start, the firm has just enough business for one typist and one PC to keep busy for a day. Say that's five documents. Now suppose the firm receives a rush order from a good customer for 10 documents tomorrow. Ideally, the firm would like to use two typists and two PCs to produce twice their normal output of five documents. However, in the short turn, the firm has fixed capital, i.e. only one PC. The table below shows the situation:

# Typists (L)	1	2	3	4	5	6	
Letters/hr (TP)	5	7	8	8	8	8	For K = 1PC
MP	5	2	1	0	0	0	

Table 7.11 Short Run Production Function for Typing

In the short run, the only variable factor is labor so the only way the firm can produce more output is by hiring additional workers. What could the second worker do? What can they contribute to the firm? Perhaps they can answer the phone, which is a major impediment to completing the typing assignment. What about a third worker? Perhaps he or she could bring coffee to the first two workers. You can see both total product and marginal product for the firm above. Now here's something to think about: At what point (e.g. after how many workers) does diminishing marginal productivity kick in, and more importantly, why?

In this example, marginal productivity starts to decline after the second worker. This is because capital is fixed. The production process for typing works best with one worker and one PC. If you add more than one typist, you get seriously diminishing marginal productivity.

Consider the long run. Suppose the firm's demand increases to 15 documents per day. What might the firm do to operate more efficiently? If demand has tripled, the firm could acquire two more PCs, which would give us a new short run production function as Table 7.4 below shows.

# Typists (L)	1	2	3	4	5	5	
Letters/hr (TP)	5	6	8	8	8	8	For K = 1PC
MP	5	2	1	0	0	0	
Letters/hr (TP)	5	10	15	17	18	18	For K = 3PC
MP	5	5	5	2	1	0	

Table 7.12 Long Run Production Function for Typing

With more capital, the firm can hire three workers before diminishing productivity comes into effect. More generally, because all factors are variable, the long run production function shows the most efficient way of producing any level of output.

7.5 | Costs in the Long Run

By the end of this section, you will be able to:

- Calculate long run total cost
- Identify economies of scale, diseconomies of scale, and constant returns to scale
- Interpret graphs of long-run average cost curves and short-run average cost curves
- Analyze cost and production in the long run and short run

The long run is the period of time when all costs are variable. The long run depends on the specifics of the firm in question—it is not a precise period of time. If you have a one-year lease on your factory, then the long run is any period longer than a year, since after a year you are no longer bound by the lease. No costs are fixed in the long run. A firm can build new factories and purchase new machinery, or it can close existing facilities. In planning for the long run, the firm will compare alternative **production technologies** (or processes).

In this context, technology refers to all alternative methods of combining inputs to produce outputs. It does not refer to a specific new invention like the tablet computer. The firm will search for the production technology that allows it to produce the desired level of output at the lowest cost. After all, lower costs lead to higher profits—at least if total revenues remain unchanged. Moreover, each firm must fear that if it does not seek out the lowest-cost methods of production, then it may lose sales to competitor firms that find a way to produce and sell for less.

Choice of Production Technology

A firm can perform many tasks with a range of combinations of labor and physical capital. For example, a firm can have human beings answering phones and taking messages, or it can invest in an automated voicemail system. A firm can hire file clerks and secretaries to manage a system of paper folders and file cabinets, or it can invest in a computerized recordkeeping system that will require fewer employees. A firm can hire workers to push supplies around a factory on rolling carts, it can invest in motorized vehicles, or it can invest in robots that carry materials without a driver. Firms often face a choice between buying a many small machines, which need a worker to run each one, or buying one larger and more expensive machine, which requires only one or two workers to operate it. In short, physical capital and labor can often substitute for each other.

Consider the example of local governments hiring a private firm to clean up public parks. Three different combinations of labor and physical capital for cleaning up a single average-sized park appear in **Table 7.13**. The first production technology is heavy on workers and light on machines, while the next two technologies substitute machines for workers. Since all three of these production methods produce the same thing—one cleaned-up park—a profit-seeking firm will choose the production technology that is least expensive, given the prices of labor and machines.

Production technology 1	10 workers	2 machines
Production technology 2	7 workers	4 machines
Production technology 3	3 workers	7 machines

Table 7.13 Three Ways to Clean a Park

Production technology 1 uses the most labor and least machinery, while production technology 3 uses the least labor and the most machinery. **Table 7.14** outlines three examples of how the total cost will change with each production technology as the cost of labor changes. As the cost of labor rises from example A to B to C, the firm will choose to substitute away from labor and use more machinery.

Example A: Workers cost $40, machines cost $80			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	10 × $40 = $400	2 × $80 = $160	$560
Cost of technology 2	7 × $40 = $280	4 × $80 = $320	$600
Cost of technology 3	3 × $40 = $120	7 × $80 = $560	$680
Example B: Workers cost $55, machines cost $80			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	10 × $55 = $550	2 × $80 = $160	$710
Cost of technology 2	7 × $55 = $385	4 × $80 = $320	$705
Cost of technology 3	3 × $55 = $165	7 × $80 = $560	$725

Table 7.14 Total Cost with Rising Labor Costs

Example C: Workers cost $90, machines cost $80

	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	10 × $90 = $900	2 × $80 = $160	$1,060
Cost of technology 2	7 × $90 = $630	4 × $80 = $320	$950
Cost of technology 3	3 × $90 = $270	7 × $80 = $560	$830

Table 7.14 Total Cost with Rising Labor Costs

Example A shows the firm's cost calculation when wages are $40 and machines costs are $80. In this case, technology 1 is the low-cost production technology. In example B, wages rise to $55, while the cost of machines does not change, in which case technology 2 is the low-cost production technology. If wages keep rising up to $90, while the cost of machines remains unchanged, then technology 3 clearly becomes the low-cost form of production, as example C shows.

This example shows that as an input becomes more expensive (in this case, the labor input), firms will attempt to conserve on using that input and will instead shift to other inputs that are relatively less expensive. This pattern helps to explain why the demand curve for labor (or any input) slopes down; that is, as labor becomes relatively more expensive, profit-seeking firms will seek to substitute the use of other inputs. When a multinational employer like Coca-Cola or McDonald's sets up a bottling plant or a restaurant in a high-wage economy like the United States, Canada, Japan, or Western Europe, it is likely to use production technologies that conserve on the number of workers and focuses more on machines. However, that same employer is likely to use production technologies with more workers and less machinery when producing in a lower-wage country like Mexico, China, or South Africa.

Economies of Scale

Once a firm has determined the least costly production technology, it can consider the optimal scale of production, or quantity of output to produce. Many industries experience economies of scale. Economies of scale refers to the situation where, as the quantity of output goes up, the cost per unit goes down. This is the idea behind "warehouse stores" like Costco or Walmart. In everyday language: a larger factory can produce at a lower average cost than a smaller factory.

Figure 7.9 illustrates the idea of economies of scale, showing the average cost of producing an alarm clock falling as the quantity of output rises. For a small-sized factory like S, with an output level of 1,000, the average cost of production is $12 per alarm clock. For a medium-sized factory like M, with an output level of 2,000, the average cost of production falls to $8 per alarm clock. For a large factory like L, with an output of 5,000, the average cost of production declines still further to $4 per alarm clock.

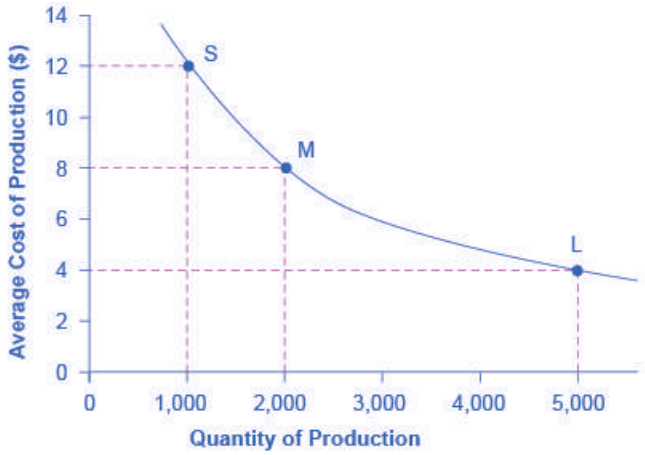


Figure 7.9 Economies of Scale A small factory like S produces 1,000 alarm clocks at an average cost of $12 per clock. A medium factory like M produces 2,000 alarm clocks at a cost of $8 per clock. A large factory like L produces 5,000 alarm clocks at a cost of $4 per clock. Economies of scale exist when the larger scale of production leads to lower average costs.

The average cost curve in **Figure 7.9** may appear similar to the average cost curves we presented earlier in this chapter, although it is downward-sloping rather than U-shaped. However, there is one major difference. The economies of scale curve is a long-run average cost curve, because it allows all factors of production to change. The short-run average cost curves we presented earlier in this chapter assumed the existence of fixed costs, and only variable costs were allowed to change.

One prominent example of economies of scale occurs in the chemical industry. Chemical plants have many pipes. The cost of the materials for producing a pipe is related to the circumference of the pipe and its length. However, the cross-section area of the pipe determines the volume of chemicals that can flow through it. The calculations in **Table 7.15** show that a pipe which uses twice as much material to make (as shown by the circumference) can actually carry four times the volume of chemicals because the pipe's cross-section area rises by a factor of four (as the Area column below shows).

	Circumference ($2\pi r$)	Area (πr^2)
4-inch pipe	12.5 inches	12.5 square inches
8-inch pipe	25.1 inches	50.2 square inches
16-inch pipe	50.2 inches	201.1 square inches

Table 7.15 Comparing Pipes: Economies of Scale in the Chemical Industry

A doubling of the cost of producing the pipe allows the chemical firm to process four times as much material. This pattern is a major reason for economies of scale in chemical production, which uses a large quantity of pipes. Of course, economies of scale in a chemical plant are more complex than this simple calculation suggests. However, the chemical engineers who design these plants have long used what they call the “six-tenths rule,” a rule of thumb which holds that increasing the quantity produced in a chemical plant by a certain percentage will increase total cost by only six-tenths as much.

Shapes of Long-Run Average Cost Curves

While in the short run firms are limited to operating on a single average cost curve (corresponding to the level of fixed costs they have chosen), in the long run when all costs are variable, they can choose to operate on any average cost curve. Thus, the **long-run average cost (LRAC) curve** is actually based on a group of **short-run average cost (SRAC) curves**, each of which represents one specific level of fixed costs. More precisely, the long-run average cost curve will be the least expensive average cost curve for any level of output. **Figure 7.10** shows how we build the

long-run average cost curve from a group of short-run average cost curves. Five short-run-average cost curves appear on the diagram. Each SRAC curve represents a different level of fixed costs. For example, you can imagine $SRAC_1$ as a small factory, $SRAC_2$ as a medium factory, $SRAC_3$ as a large factory, and $SRAC_4$ and $SRAC_5$ as very large and ultra-large. Although this diagram shows only five SRAC curves, presumably there are an infinite number of other SRAC curves between the ones that we show. Think of this family of short-run average cost curves as representing different choices for a firm that is planning its level of investment in fixed cost physical capital—knowing that different choices about capital investment in the present will cause it to end up with different short-run average cost curves in the future.

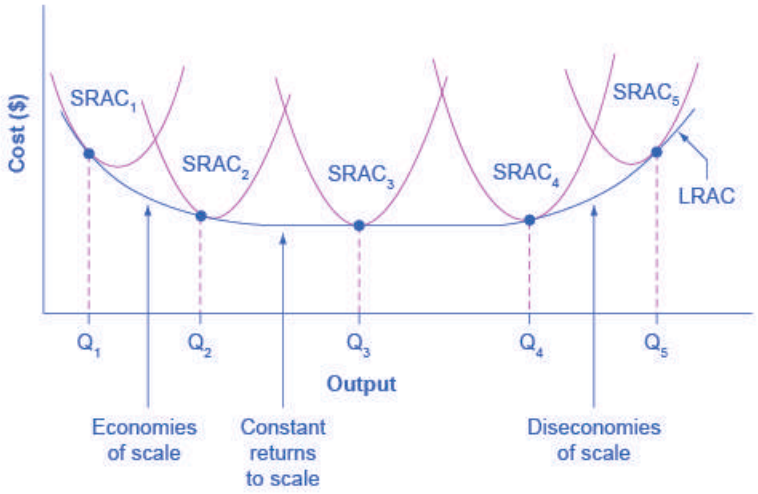


Figure 7.10 From Short-Run Average Cost Curves to Long-Run Average Cost Curves The five different short-run average cost (SRAC) curves each represents a different level of fixed costs, from the low level of fixed costs at $SRAC_1$ to the high level of fixed costs at $SRAC_5$. Other SRAC curves, not in the diagram, lie between the ones that are here. The long-run average cost (LRAC) curve shows the lowest cost for producing each quantity of output when fixed costs can vary, and so it is formed by the bottom edge of the family of SRAC curves. If a firm wished to produce quantity Q_3, it would choose the fixed costs associated with $SRAC_3$.

The long-run average cost curve shows the cost of producing each quantity in the long run, when the firm can choose its level of fixed costs and thus choose which short-run average costs it desires. If the firm plans to produce in the long run at an output of Q_3, it should make the set of investments that will lead it to locate on $SRAC_3$, which allows producing q_3 at the lowest cost. A firm that intends to produce Q_3 would be foolish to choose the level of fixed costs at $SRAC_2$ or $SRAC_4$. At $SRAC_2$ the level of fixed costs is too low for producing Q_3 at lowest possible cost, and producing q_3 would require adding a very high level of variable costs and make the average cost very high. At $SRAC_4$, the level of fixed costs is too high for producing q_3 at lowest possible cost, and again average costs would be very high as a result.

The shape of the long-run cost curve, in **Figure 7.10**, is fairly common for many industries. The left-hand portion of the long-run average cost curve, where it is downward- sloping from output levels Q_1 to Q_2 to Q_3, illustrates the case of economies of scale. In this portion of the long-run average cost curve, larger scale leads to lower average costs. We illustrated this pattern earlier in **Figure 7.9**.

In the middle portion of the long-run average cost curve, the flat portion of the curve around Q_3, economies of scale have been exhausted. In this situation, allowing all inputs to expand does not much change the average cost of production. We call this **constant returns to scale**. In this LRAC curve range, the average cost of production does not change much as scale rises or falls. The following Clear It Up feature explains where diminishing marginal returns fit into this analysis.

Clear It Up

How do economies of scale compare to diminishing marginal returns?

The concept of economies of scale, where average costs decline as production expands, might seem to conflict with the idea of diminishing marginal returns, where marginal costs rise as production expands. However, diminishing marginal returns refers only to the short-run average cost curve, where one variable input (like labor) is increasing, but other inputs (like capital) are fixed. Economies of scale refers to the long-run average cost curve where all inputs are allowed to increase together. Thus, it is quite possible and common to have an industry that has both diminishing marginal returns when only one input is allowed to change, and at the same time has economies of scale when all inputs change together to produce a larger-scale operation.

Finally, the right-hand portion of the long-run average cost curve, running from output level Q_4 to Q_5, shows a situation where, as the level of output and the scale rises, average costs rise as well. We call this situation **diseconomies of scale**. A firm or a factory can grow so large that it becomes very difficult to manage, resulting in unnecessarily high costs as many layers of management try to communicate with workers and with each other, and as failures to communicate lead to disruptions in the flow of work and materials. Not many overly large factories exist in the real world, because with their very high production costs, they are unable to compete for long against plants with lower average costs of production. However, in some planned economies, like the economy of the old Soviet Union, plants that were so large as to be grossly inefficient were able to continue operating for a long time because government economic planners protected them from competition and ensured that they would not make losses.

Diseconomies of scale can also be present across an entire firm, not just a large factory. The leviathan effect can hit firms that become too large to run efficiently, across the entirety of the enterprise. Firms that shrink their operations are often responding to finding itself in the diseconomies region, thus moving back to a lower average cost at a lower output level.

Link It Up

Visit this website (http://openstaxcollege.org/l/Toobig) to read an article about the complexity of the belief that banks can be “too-big-to-fail.”

The Size and Number of Firms in an Industry

The shape of the long-run average cost curve has implications for how many firms will compete in an industry, and whether the firms in an industry have many different sizes, or tend to be the same size. For example, say that the appliance industry sells one million dishwashers every year at a price of $500 each and the long-run average cost curve for dishwashers is in **Figure 7.11** (a). In **Figure 7.11** (a), the lowest point of the LRAC curve occurs at a quantity of 10,000 produced. Thus, the market for dishwashers will consist of 100 different manufacturing plants of this same size. If some firms built a plant that produced 5,000 dishwashers per year or 25,000 dishwashers per year, the average costs of production at such plants would be well above $500, and the firms would not be able to compete.

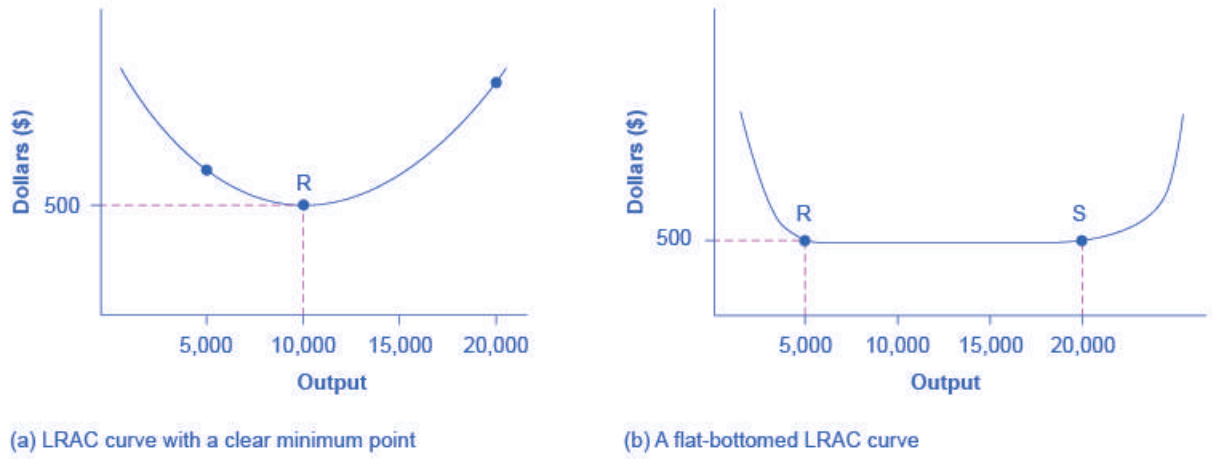


Figure 7.11 The LRAC Curve and the Size and Number of Firms (a) Low-cost firms will produce at output level R. When the LRAC curve has a clear minimum point, then any firm producing a different quantity will have higher costs. In this case, a firm producing at a quantity of 10,000 will produce at a lower average cost than a firm producing, say, 5,000 or 20,000 units. (b) Low-cost firms will produce between output levels R and S. When the LRAC curve has a flat bottom, then firms producing at any quantity along this flat bottom can compete. In this case, any firm producing a quantity between 5,000 and 20,000 can compete effectively, although firms producing less than 5,000 or more than 20,000 would face higher average costs and be unable to compete.

Clear It Up

How can we view cities as examples of economies of scale?

Why are people and economic activity concentrated in cities, rather than distributed evenly across a country? The fundamental reason must be related to the idea of economies of scale—that grouping economic activity is more productive in many cases than spreading it out. For example, cities provide a large group of nearby customers, so that businesses can produce at an efficient economy of scale. They also provide a large group of workers and suppliers, so that business can hire easily and purchase whatever specialized inputs they need. Many of the attractions of cities, like sports stadiums and museums, can operate only if they can draw on a large nearby population base. Cities are big enough to offer a wide variety of products, which is what appeals to many shoppers.

These factors are not exactly economies of scale in the narrow sense of the production function of a single firm, but they are related to growth in the overall size of population and market in an area. Cities are sometimes called “agglomeration economies.”

These agglomeration factors help to explain why every economy, as it develops, has an increasing proportion of its population living in urban areas. In the United States, about 80% of the population now lives in metropolitan areas (which include the suburbs around cities), compared to just 40% in 1900. However, in poorer nations of the world, including much of Africa, the proportion of the population in urban areas is only about 30%. One of the great challenges for these countries as their economies grow will be to manage the growth of the great cities that will arise.

If cities offer economic advantages that are a form of economies of scale, then why don’t all or most people live in one giant city? At some point, agglomeration economies must turn into diseconomies. For example, traffic congestion may reach a point where the gains from being geographically nearby are counterbalanced by how long it takes to travel. High densities of people, cars, and factories can mean more garbage and air and water pollution. Facilities like parks or museums may become overcrowded. There may be economies of scale for negative activities like crime, because high densities of people and businesses, combined with the greater impersonality of cities, make it easier for illegal activities as well as legal ones. The future of cities, both in the United States and in other countries around the world, will be determined by their ability to benefit

from the economies of agglomeration and to minimize or counterbalance the corresponding diseconomies.

We illustrate a more common case in **Figure 7.11** (b), where the LRAC curve has a flat-bottomed area of constant returns to scale. In this situation, any firm with a level of output between 5,000 and 20,000 will be able to produce at about the same level of average cost. Given that the market will demand one million dishwashers per year at a price of $500, this market might have as many as 200 producers (that is, one million dishwashers divided by firms making 5,000 each) or as few as 50 producers (one million dishwashers divided by firms making 20,000 each). The producers in this market will range in size from firms that make 5,000 units to firms that make 20,000 units. However, firms that produce below 5,000 units or more than 20,000 will be unable to compete, because their average costs will be too high. Thus, if we see an industry where almost all plants are the same size, it is likely that the long-run average cost curve has a unique bottom point as in **Figure 7.11** (a). However, if the long-run average cost curve has a wide flat bottom like **Figure 7.11** (b), then firms of a variety of different sizes will be able to compete with each other.

We can interpret the flat section of the long-run average cost curve in **Figure 7.11** (b) in two different ways. One interpretation is that a single manufacturing plant producing a quantity of 5,000 has the same average costs as a single manufacturing plant with four times as much capacity that produces a quantity of 20,000. The other interpretation is that one firm owns a single manufacturing plant that produces a quantity of 5,000, while another firm owns four separate manufacturing plants, which each produce a quantity of 5,000. This second explanation, based on the insight that a single firm may own a number of different manufacturing plants, is especially useful in explaining why the long-run average cost curve often has a large flat segment—and thus why a seemingly smaller firm may be able to compete quite well with a larger firm. At some point, however, the task of coordinating and managing many different plants raises the cost of production sharply, and the long-run average cost curve slopes up as a result.

In the examples to this point, the quantity demanded in the market is quite large (one million) compared with the quantity produced at the bottom of the long-run average cost curve (5,000, 10,000 or 20,000). In such a situation, the market is set for competition between many firms. However, what if the bottom of the long-run average cost curve is at a quantity of 10,000 and the total market demand at that price is only slightly higher than that quantity—or even somewhat lower?

Return to **Figure 7.11** (a), where the bottom of the long-run average cost curve is at 10,000, but now imagine that the total quantity of dishwashers demanded in the market at that price of $500 is only 30,000. In this situation, the total number of firms in the market would be three. We call a handful of firms in a market an “oligopoly,” and the chapter on **Monopolistic Competition and Oligopoly** will discuss the range of competitive strategies that can occur when oligopolies compete.

Alternatively, consider a situation, again in the setting of **Figure 7.11** (a), where the bottom of the long-run average cost curve is 10,000, but total demand for the product is only 5,000. (For simplicity, imagine that this demand is highly inelastic, so that it does not vary according to price.) In this situation, the market may well end up with a single firm—a monopoly—producing all 5,000 units. If any firm tried to challenge this monopoly while producing a quantity lower than 5,000 units, the prospective competitor firm would have a higher average cost, and so it would not be able to compete in the longer term without losing money. The chapter on **Monopoly** discusses the situation of a monopoly firm.

Thus, the shape of the long-run average cost curve reveals whether competitors in the market will be different sizes. If the LRAC curve has a single point at the bottom, then the firms in the market will be about the same size, but if the LRAC curve has a flat-bottomed segment of constant returns to scale, then firms in the market may be a variety of different sizes.

The relationship between the quantity at the minimum of the long-run average cost curve and the quantity demanded in the market at that price will predict how much competition is likely to exist in the market. If the quantity demanded in the market far exceeds the quantity at the minimum of the LRAC, then many firms will compete. If the quantity demanded in the market is only slightly higher than the quantity at the minimum of the LRAC, a few firms will compete. If the quantity demanded in the market is less than the quantity at the minimum of the LRAC, a single-producer monopoly is a likely outcome.

Shifting Patterns of Long-Run Average Cost

New developments in production technology can shift the long-run average cost curve in ways that can alter the size distribution of firms in an industry.

For much of the twentieth century, the most common change had been to see alterations in technology, like the assembly line or the large department store, where large-scale producers seemed to gain an advantage over smaller ones. In the long-run average cost curve, the downward-sloping economies of scale portion of the curve stretched over a larger quantity of output.

However, new production technologies do not inevitably lead to a greater average size for firms. For example, in recent years some new technologies for generating electricity on a smaller scale have appeared. The traditional coal-burning electricity plants needed to produce 300 to 600 megawatts of power to exploit economies of scale fully. However, high-efficiency turbines to produce electricity from burning natural gas can produce electricity at a competitive price while producing a smaller quantity of 100 megawatts or less. These new technologies create the possibility for smaller companies or plants to generate electricity as efficiently as large ones. Another example of a technology-driven shift to smaller plants may be taking place in the tire industry. A traditional mid-size tire plant produces about six million tires per year. However, in 2000, the Italian company Pirelli introduced a new tire factory that uses many robots. The Pirelli tire plant produced only about one million tires per year, but did so at a lower average cost than a traditional mid-sized tire plant.

Controversy has simmered in recent years over whether the new information and communications technologies will lead to a larger or smaller size for firms. On one side, the new technology may make it easier for small firms to reach out beyond their local geographic area and find customers across a state, or the nation, or even across international boundaries. This factor might seem to predict a future with a larger number of small competitors. On the other side, perhaps the new information and communications technology will create “winner-take-all” markets where one large company will tend to command a large share of total sales, as Microsoft has done producing of software for personal computers or Amazon has done in online bookselling. Moreover, improved information and communication technologies might make it easier to manage many different plants and operations across the country or around the world, and thus encourage larger firms. This ongoing battle between the forces of smallness and largeness will be of great interest to economists, businesspeople, and policymakers.

Bring it Home

Amazon

Traditionally, bookstores have operated in retail locations with inventories held either on the shelves or in the back of the store. These retail locations were very pricey in terms of rent. Until recently, Amazon had no retail locations. It only sold online and delivered by mail. Amazon now has retail stores in California, Oregon and Washington State and retail stores are coming to Illinois, Massachusetts, New Jersey, and New York. Amazon offers almost any book in print, convenient purchasing, and prompt delivery by mail. Amazon holds its inventories in huge warehouses in low-rent locations around the world. The warehouses are highly computerized using robots and relatively low-skilled workers, making for low average costs per sale. Amazon demonstrates the significant advantages economies of scale can offer to a firm that exploits those economies.

KEY TERMS

accounting profit total revenues minus explicit costs, including depreciation

average profit profit divided by the quantity of output produced; also known as profit margin

average total cost total cost divided by the quantity of output

average variable cost variable cost divided by the quantity of output

constant returns to scale expanding all inputs proportionately does not change the average cost of production

diminishing marginal productivity general rule that as a firm employs more labor, eventually the amount of additional output produced declines

diseconomies of scale the long-run average cost of producing output increases as total output increases

economic profit total revenues minus total costs (explicit plus implicit costs)

economies of scale the long-run average cost of producing output decreases as total output increases

economies of scale the long-run average cost of producing output decreases as total output increases

explicit costs out-of-pocket costs for a firm, for example, payments for wages and salaries, rent, or materials

factors of production (or inputs) resources that firms use to produce their products, for example, labor and capital

firm an organization that combines inputs of labor, capital, land, and raw or finished component materials to produce outputs.

fixed cost cost of the fixed inputs; expenditure that a firm must make before production starts and that does not change regardless of the production level

fixed inputs factors of production that can't be easily increased or decreased in a short period of time

implicit costs opportunity cost of resources already owned by the firm and used in business, for example, expanding a factory onto land already owned

long run period of time during which all of a firm's inputs are variable

long-run average cost (LRAC) curve shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology

marginal cost the additional cost of producing one more unit; mathematically, $MC = \Delta TC / \Delta L$

marginal product change in a firm's output when it employees more labor; mathematically, $MP = \Delta TP / \Delta L$

private enterprise the ownership of businesses by private individuals

production the process of combining inputs to produce outputs, ideally of a value greater than the value of the inputs

production function mathematical equation that tells how much output a firm can produce with given amounts of the inputs

production technologies alternative methods of combining inputs to produce output

revenue income from selling a firm's product; defined as price times quantity sold

short run period of time during which at least one or more of the firm's inputs is fixed

short-run average cost (SRAC) curve the average total cost curve in the short term; shows the total of the average fixed costs and the average variable costs

total cost the sum of fixed and variable costs of production

total product synonym for a firm's output

variable cost cost of production that increases with the quantity produced; the cost of the variable inputs

variable inputs factors of production that a firm can easily increase or decrease in a short period of time

KEY CONCEPTS AND SUMMARY

7.1 Explicit and Implicit Costs, and Accounting and Economic Profit

Privately owned firms are motivated to earn profits. Profit is the difference between revenues and costs. While accounting profit considers only explicit costs, economic profit considers both explicit and implicit costs.

7.2 Production in the Short Run

Production is the process a firm uses to transform inputs (e.g. labor, capital, raw materials, etc.) into outputs. It is not possible to vary fixed inputs (e.g. capital) in a short period of time. Thus, in the short run the only way to change output is to change the variable inputs (e.g. labor). Marginal product is the additional output a firm obtains by employing more labor in production. At some point, employing additional labor leads to diminishing marginal productivity, meaning the additional output obtained is less than for the previous increment to labor. Mathematically, marginal product is the slope of the total product curve.

7.3 Costs in the Short Run

For every input (e.g. labor), there is an associated factor payment (e.g. wages and salaries). The cost of production for a given quantity of output is the sum of the amount of each input required to produce that quantity of output times the associated factor payment.

In a short-run perspective, we can divide a firm's total costs into fixed costs, which a firm must incur before producing any output, and variable costs, which the firm incurs in the act of producing. Fixed costs are sunk costs; that is, because they are in the past and the firm cannot alter them, they should play no role in economic decisions about future production or pricing. Variable costs typically show diminishing marginal returns, so that the marginal cost of producing higher levels of output rises.

We calculate marginal cost by taking the change in total cost (or the change in variable cost, which will be the same thing) and dividing it by the change in output, for each possible change in output. Marginal costs are typically rising. A firm can compare marginal cost to the additional revenue it gains from selling another unit to find out whether its marginal unit is adding to profit.

We calculate average total cost by taking total cost and dividing by total output at each different level of output. Average costs are typically U-shaped on a graph. If a firm's average cost of production is lower than the market price, a firm will be earning profits.

We calculate average variable cost by taking variable cost and dividing by the total output at each level of output. Average variable costs are typically U-shaped. If a firm's average variable cost of production is lower than the market price, then the firm would be earning profits if fixed costs are left out of the picture.

7.4 Production in the Long Run

In the long run, all inputs are variable. Since diminishing marginal productivity is caused by fixed capital, there are no diminishing returns in the long run. Firms can choose the optimal capital stock to produce their desired level of output.

7.5 Costs in the Long Run

A production technology refers to a specific combination of labor, physical capital, and technology that makes up a particular method of production.

In the long run, firms can choose their production technology, and so all costs become variable costs. In making this choice, firms will try to substitute relatively inexpensive inputs for relatively expensive inputs where possible, so as to produce at the lowest possible long-run average cost.

Economies of scale refers to a situation where as the level of output increases, the average cost decreases. Constant returns to scale refers to a situation where average cost does not change as output increases. Diseconomies of scale refers to a situation where as output increases, average costs also increase.

The long-run average cost curve shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology. A downward-sloping LRAC shows economies of scale; a flat LRAC shows constant returns to scale; an upward-sloping LRAC shows diseconomies of scale. If the long-run average cost curve has only one quantity produced that results in the lowest possible average cost, then all of the firms competing in an industry should be the same size. However, if the LRAC has a flat segment at the bottom, so that a firm can produce a range of different quantities at the lowest average cost, the firms competing in the industry will display a range of sizes. The market demand in conjunction with the long-run average cost curve determines how many firms will exist in a given industry.

If the quantity demanded in the market of a certain product is much greater than the quantity found at the bottom of the long-run average cost curve, where the cost of production is lowest, the market will have many firms competing. If the quantity demanded in the market is less than the quantity at the bottom of the LRAC, there will likely be only one firm.

SELF-CHECK QUESTIONS

1. A firm had sales revenue of $1 million last year. It spent $600,000 on labor, $150,000 on capital and $200,000 on materials. What was the firm's accounting profit?

2. Continuing from **Exercise 7.1**, the firm's factory sits on land owned by the firm that it could rent for $30,000 per year. What was the firm's economic profit last year?

3. The WipeOut Ski Company manufactures skis for beginners. Fixed costs are $30. Fill in **Table 7.16** for total cost, average variable cost, average total cost, and marginal cost.

Quantity	Variable Cost	Fixed Cost	Total Cost	Average Variable Cost	Average Total Cost	Marginal Cost
0	0	$30				
1	$10	$30				
2	$25	$30				
3	$45	$30				
4	$70	$30				
5	$100	$30				
6	$135	$30				

Table 7.16

4. Based on your answers to the WipeOut Ski Company in **Exercise 7.3**, now imagine a situation where the firm produces a quantity of 5 units that it sells for a price of $25 each.

a. What will be the company's profits or losses?
b. How can you tell at a glance whether the company is making or losing money at this price by looking at average cost?
c. At the given quantity and price, is the marginal unit produced adding to profits?

5. If two painters can paint 200 square feet of wall in an hour, and three painters can paint 275 square feet, what is the marginal product of the third painter?

6. Return to the problem explained in **Table 7.13** and **Table 7.14**. If the cost of labor remains at $40, but the cost of a machine decreases to $50, what would be the total cost of each method of production? Which method should the firm use, and why?

7. Suppose the cost of machines increases to $55, while the cost of labor stays at $40. How would that affect the total cost of the three methods? Which method should the firm choose now?

8. Automobile manufacturing is an industry subject to significant economies of scale. Suppose there are four domestic auto manufacturers, but the demand for domestic autos is no more than 2.5 times the quantity produced at the bottom of the long-run average cost curve. What do you expect will happen to the domestic auto industry in the long run?

REVIEW QUESTIONS

9. What are explicit and implicit costs?

10. Would you consider an interest payment on a loan to a firm an explicit or implicit cost?

11. What is the difference between accounting and economic profit?

12. What is a production function?

13. What is the difference between a fixed input and a variable input?

14. How do we calculate marginal product?

15. What shapes would you generally expect a total product curve and a marginal product curve to have?

16. What are the factor payments for land, labor, and capital?

17. What is the difference between fixed costs and variable costs?

18. How do we calculate each of the following: marginal cost, average total cost, and average variable cost?

19. What shapes would you generally expect each of the following cost curves to have: fixed costs, variable costs, marginal costs, average total costs, and average variable costs?

20. Are there fixed costs in the long-run? Explain briefly.

21. Are fixed costs also sunk costs? Explain.

22. What are diminishing marginal returns as they relate to costs?

23. Which costs are measured on per-unit basis: fixed costs, average cost, average variable cost, variable costs, and marginal cost?

24. What is a production technology?

25. In choosing a production technology, how will firms react if one input becomes relatively more expensive?

26. What is a long-run average cost curve?

27. What is the difference between economies of scale, constant returns to scale, and diseconomies of scale?

28. What shape of a long-run average cost curve illustrates economies of scale, constant returns to scale, and diseconomies of scale?

29. Why will firms in most markets be located at or close to the bottom of the long-run average cost curve?

CRITICAL THINKING QUESTIONS

30. Small "Mom and Pop firms," like inner city grocery stores, sometimes exist even though they do not earn economic profits. How can you explain this?

31. A common name for fixed cost is "overhead." If you divide fixed cost by the quantity of output produced, you get average fixed cost. Supposed fixed cost is $1,000. What does the average fixed cost curve look like? Use your response to explain what "spreading the overhead" means.

32. How does fixed cost affect marginal cost? Why is this relationship important?

33. Average cost curves (except for average fixed cost) tend to be U-shaped, decreasing and then increasing. Marginal cost curves have the same shape, though this may be harder to see since most of the marginal cost curve is increasing. Why do you think that average and marginal cost curves have the same general shape?

34. What is the relationship between marginal product and marginal cost? (Hint: Look at the curves.) Why do you suppose that is? Is this relationship the same in the long run as in the short run?

35. It is clear that businesses operate in the short run, but do they ever operate in the long run? Discuss.

36. Return to **Table 7.2**. In the top half of the table, at what point does diminishing marginal productivity kick in? What about in the bottom half of the table? How do you explain this?

37. How would an improvement in technology, like the high-efficiency gas turbines or Pirelli tire plant, affect the long-run average cost curve of a firm? Can you draw the old curve and the new one on the same axes? How might such an improvement affect other firms in the industry?

38. Do you think that the taxicab industry in large cities would be subject to significant economies of scale? Why or why not?

PROBLEMS

39. A firm is considering an investment that will earn a 6% rate of return. If it were to borrow the money, it would have to pay 8% interest on the loan, but it currently has the cash, so it will not need to borrow. Should the firm make the investment? Show your work.

40. Return to **Figure 7.7**. What is the marginal gain in output from increasing the number of barbers from 4 to 5 and from 5 to 6? Does it continue the pattern of diminishing marginal returns?

41. Compute the average total cost, average variable cost, and marginal cost of producing 60 and 72 haircuts. Draw the graph of the three curves between 60 and 72 haircuts.

42. A small company that shovels sidewalks and driveways has 100 homes signed up for its services this winter. It can use various combinations of capital and labor: intensive labor with hand shovels, less labor with snow blowers, and still less labor with a pickup truck that has a snowplow on front. To summarize, the method choices are:

Method 1: 50 units of labor, 10 units of capital

Method 2: 20 units of labor, 40 units of capital

Method 3: 10 units of labor, 70 units of capital

If hiring labor for the winter costs $100/unit and a unit of capital costs $400, what is the best production method? What method should the company use if the cost of labor rises to $200/unit?

8 | Perfect Competition

Figure 8.1 Depending upon the competition and prices offered, a wheat farmer may choose to grow a different crop. (Credit: modification of work by Daniel X. O'Neil/Flickr Creative Commons)

Bring it Home

A Dime a Dozen

When you were younger did you babysit, deliver papers, or mow the lawn for money? If so, you faced stiff competition from many other competitors who offered identical services. There was nothing to stop others from also offering their services.

All of you charged the “going rate.” If you tried to charge more, your customers would simply buy from someone else. These conditions are very similar to the conditions agricultural growers face.

Growing a crop may be more difficult to start than a babysitting or lawn mowing service, but growers face the same fierce competition. In the grand scale of world agriculture, farmers face competition from thousands of others because they sell an identical product. After all, winter wheat is winter wheat, but if they find it hard to make money with that crop, it is relatively easy for farmers to leave the marketplace for another crop. In this case, they do not sell the family farm, they switch crops.

Take the case of the upper Midwest region of the United States—for many generations the area was called “King Wheat.” According to the United States Department of Agriculture National Agricultural Statistics Service, statistics by state, in 1997, 11.6 million acres of wheat and 780,000 acres of corn were planted in North Dakota. In the intervening 20 or so years has the mix of crops changed? Since it is relatively easy to switch crops, did farmers change what they planted in response to changes in relative crop prices? We will find out at chapter’s end.

In the meantime, let's consider the topic of this chapter—the perfectly competitive market. This is a market in

which entry and exit are relatively easy and competitors are “a dime a dozen.”

Introduction to Perfect Competition

In this chapter, you will learn about:

- Perfect Competition and Why It Matters
- How Perfectly Competitive Firms Make Output Decisions
- Entry and Exit Decisions in the Long Run
- Efficiency in Perfectly Competitive Markets

Most businesses face two realities: no one is required to buy their products, and even customers who might want those products may buy from other businesses instead. Firms that operate in perfectly competitive markets face this reality. In this chapter, you will learn how such firms make decisions about how much to produce, how much profit they make, whether to stay in business or not, and many others. Industries differ from one another in terms of how many sellers there are in a specific market, how easy or difficult it is for a new firm to enter, and the type of products that they sell. Economists refer to this as an industry's **market structure**. In this chapter, we focus on perfect competition. However, in other chapters we will examine other industry types: **Monopoly** and **Monopolistic Competition and Oligopoly**.

8.1 | Perfect Competition and Why It Matters

By the end of this section, you will be able to:

- Explain the characteristics of a perfectly competitive market
- Discuss how perfectly competitive firms react in the short run and in the long run

Firms are in **perfect competition** when the following conditions occur: (1) many firms produce identical products; (2) many buyers are available to buy the product, and many sellers are available to sell the product; (3) sellers and buyers have all relevant information to make rational decisions about the product that they are buying and selling; and (4) firms can enter and leave the market without any restrictions—in other words, there is free entry and exit into and out of the market.

A perfectly competitive firm is known as a **price taker**, because the pressure of competing firms forces it to accept the prevailing equilibrium price in the market. If a firm in a perfectly competitive market raises the price of its product by so much as a penny, it will lose all of its sales to competitors. When a wheat grower, as we discussed in the Bring It Home feature, wants to know the going price of wheat, he or she has to check on the computer or listen to the radio. Supply and demand in the entire market solely determine the market price, not the individual farmer. A perfectly competitive firm must be a very small player in the overall market, so that it can increase or decrease output without noticeably affecting the overall quantity supplied and price in the market.

A perfectly competitive market is a hypothetical extreme; however, producers in a number of industries do face many competitor firms selling highly similar goods, in which case they must often act as price takers. Economists often use agricultural markets as an example. The same crops that different farmers grow are largely interchangeable. According to the United States Department of Agriculture monthly reports, in 2015, U.S. corn farmers received an average price of $6.00 per bushel. A corn farmer who attempted to sell at $7.00 per bushel, would not have found any buyers. A perfectly competitive firm will not sell below the equilibrium price either. Why should they when they can sell all they want at the higher price? Other examples of agricultural markets that operate in close to perfectly competitive markets are small roadside produce markets and small organic farmers.

Link It Up

Visit this website (http://openstaxcollege.org/l/commodities) that reveals the current value of various commodities.

This chapter examines how profit-seeking firms decide how much to produce in perfectly competitive markets. Such firms will analyze their costs as we discussed in the chapter on **Production, Costs and Industry Structure**. In the short run, the perfectly competitive firm will seek the quantity of output where profits are highest or, if profits are not possible, where losses are lowest.

In the long run, positive economic profits will attract competition as other firms enter the market. Economic losses will cause firms to exit the market. Ultimately, perfectly competitive markets will attain long-run *equilibrium* when no new firms want to enter the market and existing firms do not want to leave the market, as economic profits have been driven down to zero.

8.2 | How Perfectly Competitive Firms Make Output Decisions

By the end of this section, you will be able to:

- Calculate profits by comparing total revenue and total cost
- Identify profits and losses with the average cost curve
- Explain the shutdown point
- Determine the price at which a firm should continue producing in the short run

A perfectly competitive firm has only one major decision to make—namely, what quantity to produce. To understand this, consider a different way of writing out the basic definition of profit:

$$\begin{aligned}\text{Profit} &= \text{Total revenue} - \text{Total cost} \\ &= (\text{Price})(\text{Quantity produced}) - (\text{Average cost})(\text{Quantity produced})\end{aligned}$$

Since a perfectly competitive firm must accept the price for its output as determined by the product's market demand and supply, it cannot choose the price it charges. This is already determined in the profit equation, and so the perfectly competitive firm can sell any number of units at exactly the same price. It implies that the firm faces a perfectly elastic demand curve for its product: buyers are willing to buy any number of units of output from the firm at the market price. When the perfectly competitive firm chooses what quantity to produce, then this quantity—along with the prices prevailing in the market for output and inputs—will determine the firm's total revenue, total costs, and ultimately, level of profits.

Determining the Highest Profit by Comparing Total Revenue and Total Cost

A perfectly competitive firm can sell as large a quantity as it wishes, as long as it accepts the prevailing market price. The formula above shows that total revenue depends on the quantity sold and the price charged. If the firm sells a higher quantity of output, then total revenue will increase. If the market price of the product increases, then total revenue also increases whatever the quantity of output sold. As an example of how a perfectly competitive firm decides what quantity to produce, consider the case of a small farmer who produces raspberries and sells them frozen for $4 per pack. Sales of one pack of raspberries will bring in $4, two packs will be $8, three packs will be $12, and

so on. If, for example, the price of frozen raspberries doubles to $8 per pack, then sales of one pack of raspberries will be $8, two packs will be $16, three packs will be $24, and so on.

Table 8.1 graphically shows total revenue and total costs for the raspberry farm, also appear in **Figure 8.2**. The horizontal axis shows the quantity of frozen raspberries produced in packs. The vertical axis shows both total revenue and total costs, measured in dollars. The total cost curve intersects with the vertical axis at a value that shows the level of fixed costs, and then slopes upward. All these cost curves follow the same characteristics as the curves that we covered in the **Production, Costs and Industry Structure** chapter.

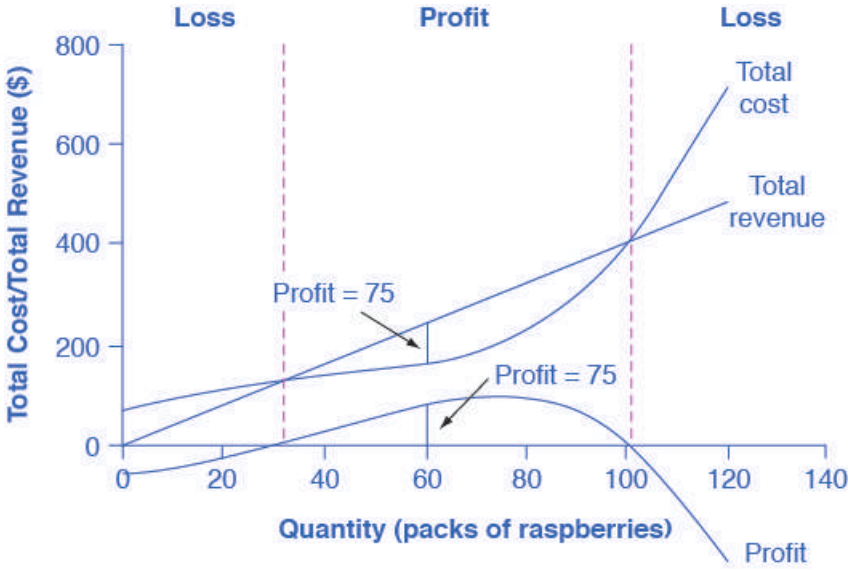


Figure 8.2 Total Cost and Total Revenue at the Raspberry Farm Total revenue for a perfectly competitive firm is a straight line sloping up. The slope is equal to the price of the good. Total cost also slopes up, but with some curvature. At higher levels of output, total cost begins to slope upward more steeply because of diminishing marginal returns. The maximum profit will occur at the quantity where the difference between total revenue and total cost is largest.

Quantity (Q)	Total Cost (TC)	Total Revenue (TR)	Profit
0	$62	$0	−$62
10	$90	$40	−$50
20	$110	$80	−$30
30	$126	$120	−$6
40	$138	$160	$22
50	$150	$200	$50
60	$165	$240	$75
70	$190	$280	$90
80	$230	$320	$90
90	$296	$360	$64

Table 8.1 Total Cost and Total Revenue at the Raspberry Farm

Quantity (Q)	Total Cost (TC)	Total Revenue (TR)	Profit
100	$400	$400	$0
110	$550	$440	$−110
120	$715	$480	$−235

Table 8.1 Total Cost and Total Revenue at the Raspberry Farm

Based on its total revenue and total cost curves, a perfectly competitive firm like the raspberry farm can calculate the quantity of output that will provide the highest level of profit. At any given quantity, total revenue minus total cost will equal profit. One way to determine the most profitable quantity to produce is to see at what quantity total revenue exceeds total cost by the largest amount. **Figure 8.2** shows total revenue, total cost and profit using the data from **Table 8.1**. The vertical gap between total revenue and total cost is profit, for example, at Q = 60, TR = 240 and TC = 165. The difference is 75, which is the height of the profit curve at that output level. The firm doesn't make a profit at every level of output. In this example, total costs will exceed total revenues at output levels from 0 to approximately 30, and so over this range of output, the firm will be making losses. At output levels from 40 to 100, total revenues exceed total costs, so the firm is earning profits. However, at any output greater than 100, total costs again exceed total revenues and the firm is making increasing losses. Total profits appear in the final column of **Table 8.1**. Maximum profit occurs at an output between 70 and 80, when profit equals $90.

A higher price would mean that total revenue would be higher for every quantity sold. A lower price would mean that total revenue would be lower for every quantity sold. What happens if the price drops low enough so that the total revenue line is completely below the total cost curve; that is, at every level of output, total costs are higher than total revenues? In this instance, the best the firm can do is to suffer losses. However, a profit-maximizing firm will prefer the quantity of output where total revenues come closest to total costs and thus where the losses are smallest.

(Later we will see that sometimes it will make sense for the firm to close, rather than stay in operation producing output.)

Comparing Marginal Revenue and Marginal Costs

The approach that we described in the previous section, using total revenue and total cost, is not the only approach to determining the profit maximizing level of output. In this section, we provide an alternative approach which uses marginal revenue and marginal cost.

Firms often do not have the necessary data they need to draw a complete total cost curve for all levels of production. They cannot be sure of what total costs would look like if they, say, doubled production or cut production in half, because they have not tried it. Instead, firms experiment. They produce a slightly greater or lower quantity and observe how it affects profits. In economic terms, this practical approach to maximizing profits means examining how changes in production affect marginal revenue and marginal cost.

Figure 8.3 presents the marginal revenue and marginal cost curves based on the total revenue and total cost in **Table 8.1**. The **marginal revenue** curve shows the additional revenue gained from selling one more unit. As mentioned before, a firm in perfect competition faces a perfectly elastic demand curve for its product—that is, the firm's demand curve is a horizontal line drawn at the market price level. This also means that the firm's marginal revenue curve is the same as the firm's demand curve: Every time a consumer demands one more unit, the firm sells one more unit and revenue increases by exactly the same amount equal to the market price. In this example, every time the firm sells a pack of frozen raspberries, the firm's revenue increases by $4. **Table 8.2** shows an example of this. This condition only holds for price taking firms in perfect competition where:

$$\text{marginal revenue} = \text{price}$$

The formula for marginal revenue is:

$$\text{marginal revenue} = \frac{\text{change in total revenue}}{\text{change in quantity}}$$

Price	Quantity	Total Revenue	Marginal Revenue
$4	1	$4	-
$4	2	$8	$4
$4	3	$12	$4
$4	4	$16	$4

Table 8.2

Notice that marginal revenue does not change as the firm produces more output. That is because under perfect competition, the price is determined through the interaction of supply and demand in the market and does not change as the farmer produces more (keeping in mind that, due to the relative small size of each firm, increasing their supply has no impact on the total market supply where price is determined).

Since a perfectly competitive firm is a price taker, it can sell whatever quantity it wishes at the market-determined price. We calculate marginal cost, the cost per additional unit sold, by dividing the change in total cost by the change in quantity. The formula for marginal cost is:

$$\text{marginal cost} = \frac{\text{change in total cost}}{\text{change in quantity}}$$

Ordinarily, marginal cost changes as the firm produces a greater quantity.

In the raspberry farm example, in **Figure 8.3**, **Figure 8.4** and **Table 8.3**, marginal cost at first declines as production increases from 10 to 20 to 30 to 40 packs of raspberries—which represents the area of increasing marginal returns that is not uncommon at low levels of production. At some point, though, marginal costs start to increase, displaying the typical pattern of diminishing marginal returns. If the firm is producing at a quantity where MR > MC, like 40 or 50 packs of raspberries, then it can increase profit by increasing output because the marginal revenue is exceeding the marginal cost. If the firm is producing at a quantity where MC > MR, like 90 or 100 packs, then it can increase profit by reducing output because the reductions in marginal cost will exceed the reductions in marginal revenue. The firm's profit-maximizing choice of output will occur where MR = MC (or at a choice close to that point).

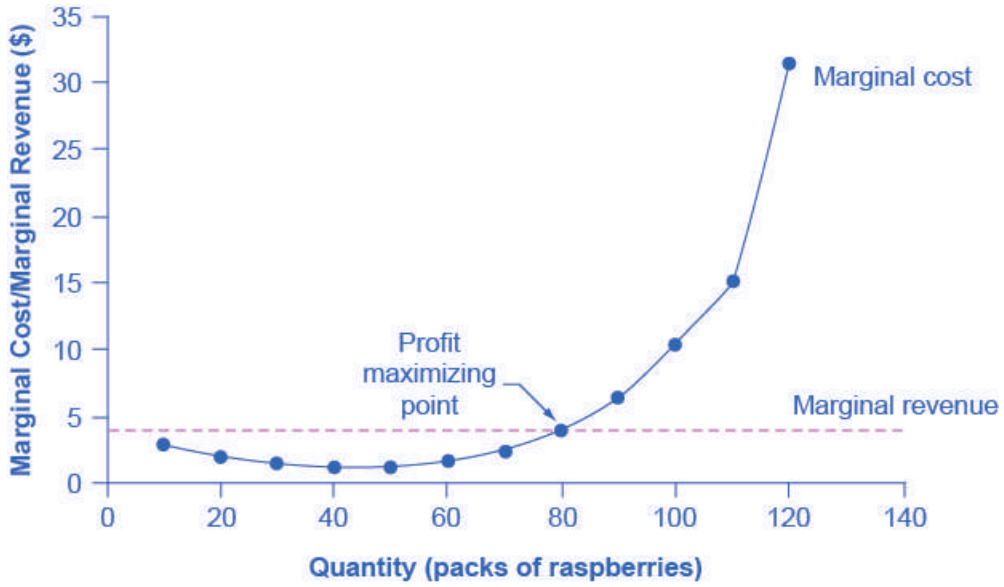


Figure 8.3 Marginal Revenues and Marginal Costs at the Raspberry Farm: Individual Farmer For a perfectly competitive firm, the marginal revenue (MR) curve is a horizontal line because it is equal to the price of the good, which is determined by the market, as **Figure 8.4** illustrates. The marginal cost (MC) curve is sometimes initially downward-sloping, if there is a region of increasing marginal returns at low levels of output, but is eventually upward-sloping at higher levels of output as diminishing marginal returns kick in.

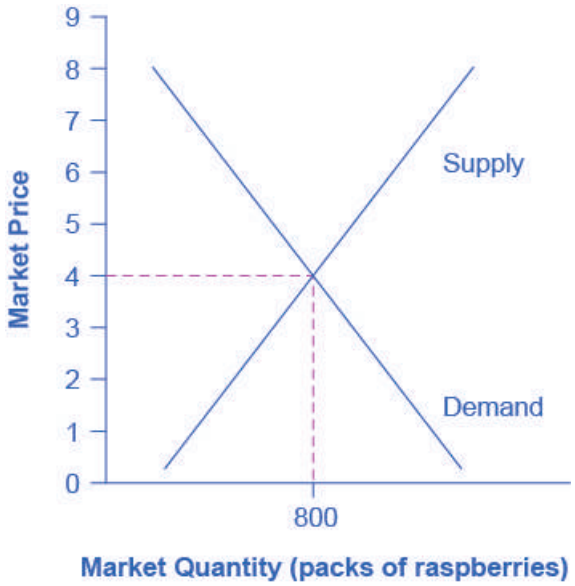


Figure 8.4 Marginal Revenues and Marginal Costs at the Raspberry Farm: Raspberry Market The equilibrium price of raspberries is determined through the interaction of market supply and market demand at $4.00.

Quantity	Total Cost	Marginal Cost	Total Revenue	Marginal Revenue	Profit
0	$62	-	$0	$4	-$62
10	$90	$2.80	$40	$4	-$50
20	$110	$2.00	$80	$4	-$30

Table 8.3 Marginal Revenues and Marginal Costs at the Raspberry Farm

Quantity	Total Cost	Marginal Cost	Total Revenue	Marginal Revenue	Profit
30	$126	$1.60	$120	$4	-$6
40	$138	$1.20	$160	$4	$22
50	$150	$1.20	$200	$4	$50
60	$165	$1.50	$240	$4	$75
70	$190	$2.50	$280	$4	$90
80	$230	$4.00	$320	$4	$90
90	$296	$6.60	$360	$4	$64
100	$400	$10.40	$400	$4	$0
110	$550	$15.00	$440	$4	-$110
120	$715	$16.50	$480	$4	-$235

Table 8.3 Marginal Revenues and Marginal Costs at the Raspberry Farm

In this example, the marginal revenue and marginal cost curves cross at a price of $4 and a quantity of 80 produced. If the farmer started out producing at a level of 60, and then experimented with increasing production to 70, marginal revenues from the increase in production would exceed marginal costs—and so profits would rise. The farmer has an incentive to keep producing. At a level of output of 80, marginal cost and marginal revenue are equal so profit doesn't change. If the farmer then experimented further with increasing production from 80 to 90, he would find that marginal costs from the increase in production are greater than marginal revenues, and so profits would decline.

The profit-maximizing choice for a perfectly competitive firm will occur at the level of output where marginal revenue is equal to marginal cost—that is, where MR = MC. This occurs at Q = 80 in the figure.

Work It Out

Does Profit Maximization Occur at a Range of Output or a Specific Level of Output?

Table 8.1 shows that maximum profit occurs at any output level between 70 and 80 units of output. But MR = MC occurs only at 80 units of output. How can do we explain this slight discrepancy? As long as MR > MC. a profit-seeking firm should keep expanding production. Expanding production into the zone where MR < MC reduces economic profits. It's true that profit is the same at Q = 70 and Q = 80, but it's only when the firm goes beyond that that see that profits fall. Thus, MR = MC is the signal to stop expanding, so that is the level of output they should target.

Because the marginal revenue received by a perfectly competitive firm is equal to the price P, we can also write the profit-maximizing rule for a perfectly competitive firm as a recommendation to produce at the quantity of output where P = MC.

Profits and Losses with the Average Cost Curve

Does maximizing profit (producing where MR = MC) imply an actual economic profit? The answer depends on the relationship between price and average total cost, which is the average profit or **profit margin**. If the market price is higher than the firm's average cost of production for that quantity produced, then the profit margin is positive and the firm will earn profits. Conversely, if the market price is lower than the average cost of production, the profit margin is negative and the firm will suffer losses. You might think that, in this situation, the firm may want to shut

down immediately. Remember, however, that the firm has already paid for fixed costs, such as equipment, so it may continue to produce for a while and incur a loss. **Table 8.3** continues the raspberry farm example. **Figure 8.5** illustrates the three possible scenarios: (a) where price intersects marginal cost at a level above the average cost curve, (b) where price intersects marginal cost at a level equal to the average cost curve, and (c) where price intersects marginal cost at a level below the average cost curve.

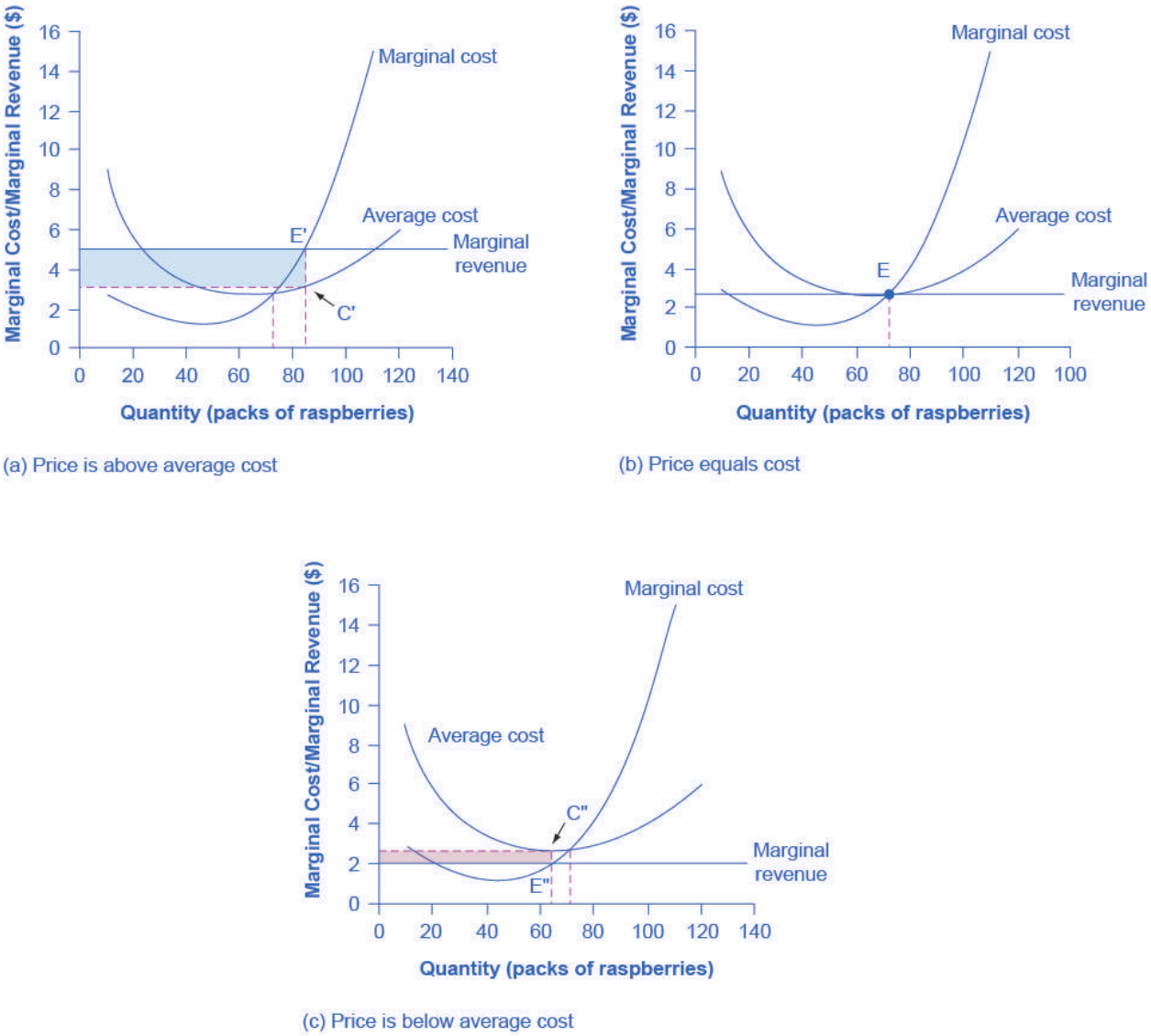


Figure 8.5 Price and Average Cost at the Raspberry Farm In (a), price intersects marginal cost above the average cost curve. Since price is greater than average cost, the firm is making a profit. In (b), price intersects marginal cost at the minimum point of the average cost curve. Since price is equal to average cost, the firm is breaking even. In (c), price intersects marginal cost below the average cost curve. Since price is less than average cost, the firm is making a loss.

First consider a situation where the price is equal to $5 for a pack of frozen raspberries. The rule for a profit-maximizing perfectly competitive firm is to produce the level of output where Price= MR = MC, so the raspberry farmer will produce a quantity of approximately 85, which is labeled as E' in **Figure 8.5** (a). Remember that the area of a rectangle is equal to its base multiplied by its height. The farm's total revenue at this price will be shown by the rectangle from the origin over to a quantity of 85 packs (the base) up to point E' (the height), over to the price of $5, and back to the origin. The average cost of producing 80 packs is shown by point C or about $3.50. Total costs will be the quantity of 85 times the average cost of $3.50, which is shown by the area of the rectangle from the origin to a quantity of 90, up to point C, over to the vertical axis and down to the origin. The difference between total revenues and total costs is profits. Thus, profits will be the blue shaded rectangle on top.

We calculate this as:

$$\begin{aligned} \text{profit} &= \text{total revenue} - \text{total cost} \\ &= (85)(\$5.00) - (85)(\$3.50) \\ &= \$170 \end{aligned}$$

Or, we can calculate it as:

$$\begin{aligned} \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\ &= (\$5.00 - \$3.50) \times 85 \\ &= \$170 \end{aligned}$$

Now consider **Figure 8.5** (b), where the price has fallen to $2.75 for a pack of frozen raspberries. Again, the perfectly competitive firm will choose the level of output where Price = MR = MC, but in this case, the quantity produced will be 75. At this price and output level, where the marginal cost curve is crossing the average cost curve, the price the firm receives is exactly equal to its average cost of production. We call this the **break even point**.

The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 75 packs (the base) up to point E (the height), over to the price of $2.75, and back to the origin. The height of the average cost curve at Q = 75, i.e. point E, shows the average cost of producing this quantity. Total costs will be the quantity of 75 times the average cost of $2.75, which is shown by the area of the rectangle from the origin to a quantity of 75, up to point E, over to the vertical axis and down to the origin. It should be clear that the rectangles for total revenue and total cost are the same. Thus, the firm is making zero profit. The calculations are as follows:

$$\begin{aligned} \text{profit} &= \text{total revenue} - \text{total cost} \\ &= (75)(\$2.75) - (75)(\$2.75) \\ &= \$0 \end{aligned}$$

Or, we can calculate it as:

$$\begin{aligned} \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\ &= (\$2.75 - \$2.75) \times 75 \\ &= \$0 \end{aligned}$$

In **Figure 8.5** (c), the market price has fallen still further to $2.00 for a pack of frozen raspberries. At this price, marginal revenue intersects marginal cost at a quantity of 65. The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 65 packs (the base) up to point E" (the height), over to the price of $2, and back to the origin. The average cost of producing 65 packs is shown by Point C" or shows the average cost of producing 50 packs is about $2.73. Total costs will be the quantity of 65 times the average cost of $2.73, which the area of the rectangle from the origin to a quantity of 50, up to point C", over to the vertical axis and down to the origin shows. It should be clear from examining the two rectangles that total revenue is less than total cost. Thus, the firm is losing money and the loss (or negative profit) will be the rose-shaded rectangle.

The calculations are:

$$\begin{aligned} \text{profit} &= (\text{total revenue} - \text{total cost}) \\ &= (65)(\$2.00) - (65)(\$2.73) \\ &= -\$47.45 \end{aligned}$$

Or:

$$\begin{aligned} \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\ &= (\$2.00 - \$2.73) \times 65 \\ &= -\$47.45 \end{aligned}$$

If the market price that perfectly competitive firm receives leads it to produce at a quantity where the price is greater than average cost, the firm will earn profits. If the price the firm receives causes it to produce at a quantity where price equals average cost, which occurs at the minimum point of the AC curve, then the firm earns zero profits. Finally, if the price the firm receives leads it to produce at a quantity where the price is less than average cost, the firm will earn losses. **Table 8.4** summarizes this.

If...	Then...
Price > ATC	Firm earns an economic profit
Price = ATC	Firm earns zero economic profit
Price < ATC	Firm earns a loss

Table 8.4

Which intersection should a firm choose?

At a price of $2, MR intersects MC at two points: Q = 20 and Q = 65. It never makes sense for a firm to choose a level of output on the downward sloping part of the MC curve, because the profit is lower (the loss is bigger). Thus, the correct choice of output is Q = 65.

The Shutdown Point

The possibility that a firm may earn losses raises a question: Why can the firm not avoid losses by shutting down and not producing at all? The answer is that shutting down can reduce variable costs to zero, but in the short run, the firm has already paid for fixed costs. As a result, if the firm produces a quantity of zero, it would still make losses because it would still need to pay for its fixed costs. Therefore when a firm is experiencing losses, it must face a question: should it continue producing or should it shut down?

As an example, consider the situation of the Yoga Center, which has signed a contract to rent space that costs $10,000 per month. If the firm decides to operate, its marginal costs for hiring yoga teachers is $15,000 for the month. If the firm shuts down, it must still pay the rent, but it would not need to hire labor. **Table 8.5** shows three possible scenarios. In the first scenario, the Yoga Center does not have any clients, and therefore does not make any revenues, in which case it faces losses of $10,000 equal to the fixed costs. In the second scenario, the Yoga Center has clients that earn the center revenues of $10,000 for the month, but ultimately experiences losses of $15,000 due to having to hire yoga instructors to cover the classes. In the third scenario, the Yoga Center earns revenues of $20,000 for the month, but experiences losses of $5,000.

In all three cases, the Yoga Center loses money. In all three cases, when the rental contract expires in the long run, assuming revenues do not improve, the firm should exit this business. In the short run, though, the decision varies depending on the level of losses and whether the firm can cover its variable costs. In scenario 1, the center does not have any revenues, so hiring yoga teachers would increase variable costs and losses, so it should shut down and only incur its fixed costs. In scenario 2, the center's losses are greater because it does not make enough revenue to offset the increased variable costs, so it should shut down immediately and only incur its fixed costs. If price is below the minimum average variable cost, the firm must shut down. In contrast, in scenario 3 the revenue that the center can earn is high enough that the losses diminish when it remains open, so the center should remain open in the short run.

Scenario 1
If the center shuts down now, revenues are zero but it will not incur any variable costs and would only need to pay fixed costs of $10,000.

Table 8.5 Should the Yoga Center Shut Down Now or Later?

$$\begin{aligned}\text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost})\\ &= 0 - \$10{,}000\\ &= -\$10{,}000\end{aligned}$$

Scenario 2
The center earns revenues of $10,000, and variable costs are $15,000. The center should shut down now.

$$\begin{aligned}\text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost})\\ &= \$10{,}000 - (\$10{,}000 + \$15{,}000)\\ &= -\$15{,}000\end{aligned}$$

Scenario 3
The center earns revenues of $20,000, and variable costs are $15,000. The center should continue in business.

$$\begin{aligned}\text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost})\\ &= \$20{,}000 - (\$10{,}000 + \$15{,}000)\\ &= -\$5{,}000\end{aligned}$$

Table 8.5 Should the Yoga Center Shut Down Now or Later?

Figure 8.6 illustrates the lesson that remaining open requires the price to exceed the firm's average variable cost. When the firm is operating below the break-even point, where price equals average cost, it is operating at a loss so it faces two options: continue to produce and lose money or shutdown. Which option is preferable? The one that loses the least money is the best choice.

At a price of $2.00 per pack, as Figure 8.6 (a) illustrates, if the farm stays in operation it will produce at a level of 65 packs of raspberries, and it will make losses of $47.45 (as explained earlier). The alternative would be to shutdown and lose all the fixed costs of $62.00. Since losing $47.45 is preferable to losing $62.00, the profit maximizing (or in this case the loss minimizing) choice is to stay in operation. The key reason is because price is above average variable cost. This means that at the current price the farm can pay all its variable costs, and have some revenue left over to pay some of the fixed costs. So the loss represents the part of the fixed costs the farm can't pay, which is less than the entire fixed costs. However, if the price declined to $1.50 per pack, as **Figure 8.6** shows (b), and if the firm applied its rule of producing where P = MR = MC, it would produce a quantity of 60. This price is below average variable cost for this level of output. If the farmer cannot pay workers (the variable costs), then it has to shut down. At this price and output, total revenues would be $90 (quantity of 60 times price of $1.50) and total cost would be $165, for overall losses of $75. If the farm shuts down, it must pay only its fixed costs of $62, so shutting down is preferable to selling at a price of $1.50 per pack.

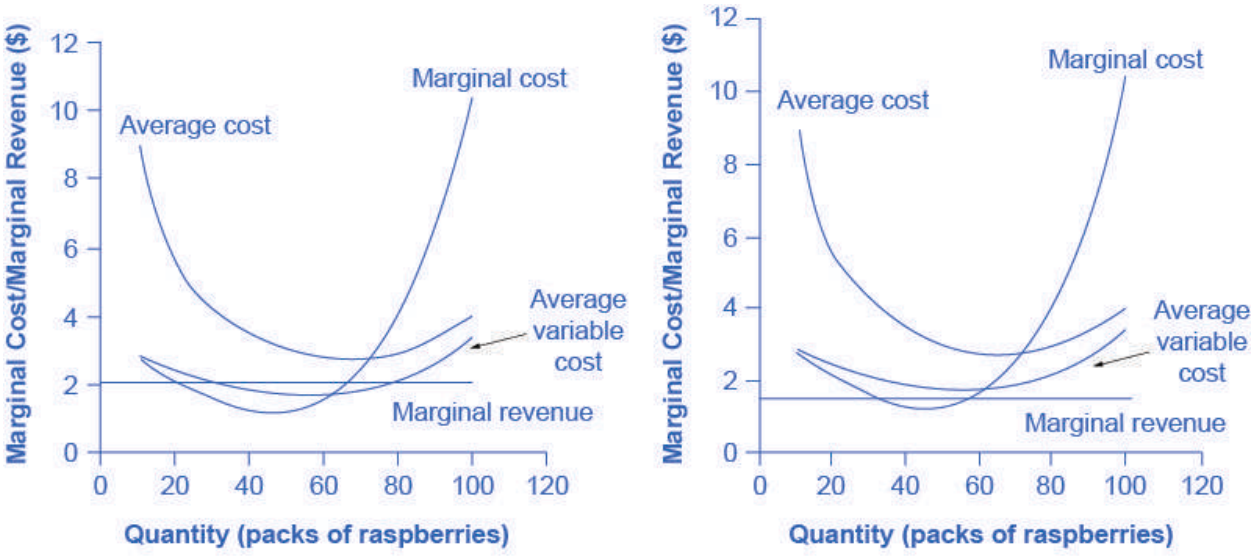


Figure 8.6 The Shutdown Point for the Raspberry Farm In (a), the farm produces at a level of 65. It is making losses of $47.50, but price is above average variable cost, so it continues to operate. In (b), total revenues are $90 and total cost is $165, for overall losses of $75. If the farm shuts down, it must pay only its fixed costs of $62. Shutting down is preferable to selling at a price of $1.50 per pack.

Looking at **Table 8.6**, if the price falls below about $1.65, the minimum average variable cost, the firm must shut down.

Quantity Q	Average Variable Cost AVC	Average Cost AC	Marginal Cost MC
0	-	-	-
10	$2.80	$9.00	$2.80
20	$2.40	$5.50	$2.00
30	$2.13	$4.20	$1.60
40	$1.90	$3.45	$1.20
50	$1.76	$3.00	$1.20
60	$1.72	$2.75	$1.50
70	$1.83	$2.71	$2.50
80	$2.10	$2.88	$4.00
90	$2.60	$3.29	$6.60
100	$3.38	$4.00	$10.40
110	$4.44	$5.00	$15.00
120	$5.44	$5.96	$31.50

Table 8.6 Cost of Production for the Raspberry Farm

The intersection of the average variable cost curve and the marginal cost curve, which shows the price below which the firm would lack enough revenue to cover its variable costs, is called the **shutdown point**. If the perfectly competitive firm faces a market price above the shutdown point, then the firm is at least covering its average variable costs. At a price above the shutdown point, the firm is also making enough revenue to cover at least a portion of fixed costs, so it should limp ahead even if it is making losses in the short run, since at least those losses will be smaller than if the firm shuts down immediately and incurs a loss equal to total fixed costs. However, if the firm is receiving a price below the price at the shutdown point, then the firm is not even covering its variable costs. In this case, staying open is making the firm's losses larger, and it should shut down immediately. To summarize, if:

- price < minimum average variable cost, then firm shuts down
- price > minimum average variable cost, then firm stays in business

Short-Run Outcomes for Perfectly Competitive Firms

The average cost and average variable cost curves divide the marginal cost curve into three segments, as **Figure 8.7** shows. At the market price, which the perfectly competitive firm accepts as given, the profit-maximizing firm chooses the output level where price or marginal revenue, which are the same thing for a perfectly competitive firm, is equal to marginal cost: P = MR = MC.

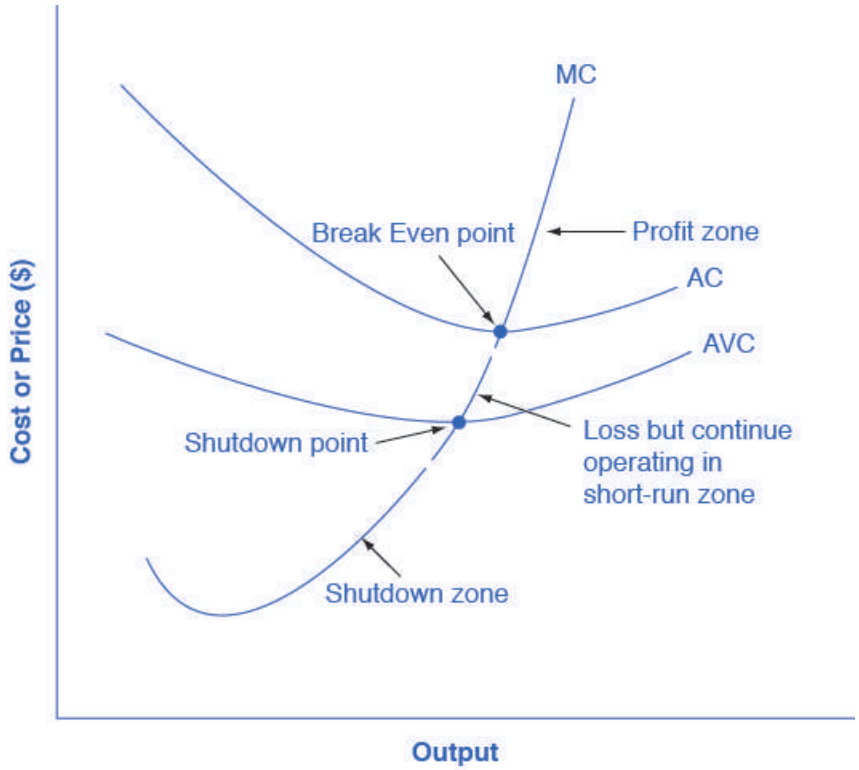


Figure 8.7 Profit, Loss, Shutdown We can divide marginal cost curve into three zones, based on where it is crossed by the average cost and average variable cost curves. We call the point where MC crosses AC the break even point. If the firm is operating where the market price is at a level higher than the break even point, then price will be greater than average cost and the firm is earning profits. If the price is exactly at the break even point, then the firm is making zero profits. If price falls in the zone between the shutdown point and the break even point, then the firm is making losses but will continue to operate in the short run, since it is covering its variable costs, and more if price is above the shutdown-point price. However, if price falls below the price at the shutdown point, then the firm will shut down immediately, since it is not even covering its variable costs.

First consider the upper zone, where prices are above the level where marginal cost (MC) crosses average cost (AC) at the zero profit point. At any price above that level, the firm will earn profits in the short run. If the price falls exactly on the break even point where the MC and AC curves cross, then the firm earns zero profits. If a price falls into the zone between the break even point, where MC crosses AC, and the shutdown point, where MC crosses AVC, the firm will be making losses in the short run—but since the firm is more than covering its variable costs, the losses

are smaller than if the firm shut down immediately. Finally, consider a price at or below the shutdown point where MC crosses AVC. At any price like this one, the firm will shut down immediately, because it cannot even cover its variable costs.

Marginal Cost and the Firm's Supply Curve

For a perfectly competitive firm, the marginal cost curve is identical to the firm's supply curve starting from the minimum point on the average variable cost curve. To understand why this perhaps surprising insight holds true, first think about what the supply curve means. A firm checks the market price and then looks at its supply curve to decide what quantity to produce. Now, think about what it means to say that a firm will maximize its profits by producing at the quantity where P = MC. This rule means that the firm checks the market price, and then looks at its marginal cost to determine the quantity to produce—and makes sure that the price is greater than the minimum average variable cost. In other words, the marginal cost curve above the minimum point on the average variable cost curve becomes the firm's supply curve.

Link It Up

Watch this video (http://openstaxcollege.org/l/foodprice) that addresses how drought in the United States can impact food prices across the world. (Note that the story on the drought is the second one in the news report. You need to let the video play through the first story in order to watch the story on the drought.)

As we discussed in the chapter on **Demand and Supply**, many of the reasons that supply curves shift relate to underlying changes in costs. For example, a lower price of key inputs or new technologies that reduce production costs cause supply to shift to the right. In contrast, bad weather or added government regulations can add to costs of certain goods in a way that causes supply to shift to the left. We can also interpret these shifts in the firm's supply curve as shifts of the marginal cost curve. A shift in costs of production that increases marginal costs at all levels of output—and shifts MC upward and to the left—will cause a perfectly competitive firm to produce less at any given market price. Conversely, a shift in costs of production that decreases marginal costs at all levels of output will shift MC downward and to the right and as a result, a competitive firm will choose to expand its level of output at any given price. The following Work It Out feature will walk you through an example.

Work It Out

At What Price Should the Firm Continue Producing in the Short Run?

To determine the short-run economic condition of a firm in perfect competition, follow the steps outlined below. Use the data in **Table 8.7**.

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	$28	$20	$0	-	-	-	-	-	-
1	$28	$20	$20	-	-	-	-	-	-
2	$28	$20	$25	-	-	-	-	-	-
3	$28	$20	$35	-	-	-	-	-	-
4	$28	$20	$52	-	-	-	-	-	-
5	$28	$20	$80	-	-	-	-	-	-

Table 8.7

Step 1. Determine the cost structure for the firm. For a given total fixed costs and variable costs, calculate total cost, average variable cost, average total cost, and marginal cost. Follow the formulas given in the **Production, Costs, and Industry Structure** chapter. These calculations are in **Table 8.8**.

Q	P	TFC	TVC	TC (TFC+TVC)	AVC (TVC/Q)	ATC (TC/Q)	MC $(TC_2-TC_1)/(Q_2-Q_1)$
0	$28	$20	$0	$20+$0=$20	-	-	-
1	$28	$20	$20	$20+$20=$40	$20/1=$20.00	$40/1=$40.00	($40−$20)/(1−0)= $20
2	$28	$20	$25	$20+$25=$45	$25/2=$12.50	$45/2=$22.50	($45−$40)/(2−1)= $5
3	$28	$20	$35	$20+$35=$55	$35/3=$11.67	$55/3=$18.33	($55−$45)/(3−2)= $10
4	$28	$20	$52	$20+$52=$72	$52/4=$13.00	$72/4=$18.00	($72−$55)/(4−3)= $17
5	$28	$20	$80	$20+$80=$100	$80/5=$16.00	$100/5=$20.00	($100−$72)/(5−4)= $28

Table 8.8

Step 2. Determine the market price that the firm receives for its product. Since the firm in perfect competition is a price taker, the market price is constant With the given price, calculate total revenue as equal to price multiplied by quantity for all output levels produced. In this example, the given price is $28. You can see that in the second column of **Table 8.9**.

Quantity	Price	Total Revenue (P × Q)
0	$28	$28×0=$0

Table 8.9

Quantity	Price	Total Revenue (P × Q)
1	$28	$28×1=$28
2	$28	$28×2=$56
3	$28	$28×3=$84
4	$28	$28×4=$112
5	$28	$28×5=$140

Table 8.9

Step 3. Calculate profits as total cost subtracted from total revenue, as Table 8.10 shows.

Quantity	Total Revenue	Total Cost	Profits (TR−TC)
0	$0	$20	$0−$20=−$20
1	$28	$40	$28−$40=−$12
2	$56	$45	$56−$45=$11
3	$84	$55	$84−$55=$29
4	$112	$72	$112−$72=$40
5	$140	$100	$140−$100=$40

Table 8.10

Step 4. To find the profit-maximizing output level, look at the Marginal Cost column (at every output level produced), as Table 8.11 shows, and determine where it is equal to the market price. The output level where price equals the marginal cost is the output level that maximizes profits.

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	$28	$20	$0	$20	-	-	-	$0	−$20
1	$28	$20	$20	$40	$20.00	$40.00	$20	$28	−$12
2	$28	$20	$25	$45	$12.50	$22.50	$5	$56	$11
3	$28	$20	$35	$55	$11.67	$18.33	$10	$84	$29
4	$28	$20	$52	$72	$13.00	$18.00	$17	$112	$40
5	**$28**	$20	$80	$100	$16.40	$20.40	**$28**	$140	**$40**

Table 8.11

Step 5. Once you have determined the profit-maximizing output level (in this case, output quantity 5), you can look at the amount of profits made (in this case, $40).

Step 6. If the firm is making economic losses, the firm needs to determine whether it produces the output level where price equals marginal revenue and equals marginal cost or it shuts down and only incurs its fixed costs.

Step 7. For the output level where marginal revenue is equal to marginal cost, check if the market price is greater than the average variable cost of producing that output level.

- If P > AVC but P < ATC, then the firm continues to produce in the short-run, making economic losses.
- If P < AVC, then the firm stops producing and only incurs its fixed costs.

In this example, the price of $28 is greater than the AVC ($16.40) of producing 5 units of output, so the firm continues producing.

8.3 | Entry and Exit Decisions in the Long Run

By the end of this section, you will be able to:

- Explain how entry and exit lead to zero profits in the long run
- Discuss the long-run adjustment process

It is impossible to precisely define the line between the short run and the long run with a stopwatch, or even with a calendar. It varies according to the specific business. Therefore, the distinction between the short run and the long run is more technical: in the short run, firms cannot change the usage of fixed inputs, while in the long run, the firm can adjust all factors of production.

In a competitive market, profits are a red cape that incites businesses to charge. If a business is making a profit in the short run, it has an incentive to expand existing factories or to build new ones. New firms may start production, as well. When new firms enter the industry in response to increased industry profits it is called **entry**.

Losses are the black thundercloud that causes businesses to flee. If a business is making losses in the short run, it will either keep limping along or just shut down, depending on whether its revenues are covering its variable costs. But in the long run, firms that are facing losses will cease production altogether. The long-run process of reducing production in response to a sustained pattern of losses is called **exit**. The following Clear It Up feature discusses where some of these losses might come from, and the reasons why some firms go out of business.

Clear It Up

Why do firms cease to exist?

Can we say anything about what causes a firm to exit an industry? Profits are the measurement that determines whether a business stays operating or not. Individuals start businesses with the purpose of making profits. They invest their money, time, effort, and many other resources to produce and sell something that they hope will give them something in return. Unfortunately, not all businesses are successful, and many new startups soon realize that their “business venture” must eventually end.

In the model of perfectly competitive firms, those that consistently cannot make money will “exit,” which is a nice, bloodless word for a more painful process. When a business fails, after all, workers lose their jobs, investors lose their money, and owners and managers can lose their dreams. Many businesses fail. The U.S. Small Business Administration indicates that in 2011, 534,907 new firms "entered," and 575,691 firms failed.

Sometimes a business fails because of poor management or workers who are not very productive, or because of tough domestic or foreign competition. Businesses also fail from a variety of causes. For example, conditions of demand and supply in the market may shift in an unexpected way, so that the prices that a business charges for outputs fall or the prices for inputs rise. With millions of businesses in the U.S. economy, even a small fraction of them failing will affect many people—and business failures can be very hard on the workers and managers directly involved. However, from the standpoint of the overall economic system, business exits are sometimes a necessary evil if a market-oriented system is going to offer a flexible mechanism for satisfying customers, keeping costs low, and inventing new products.

How Entry and Exit Lead to Zero Profits in the Long Run

No perfectly competitive firm acting alone can affect the market price. However, the combination of many firms entering or exiting the market will affect overall supply in the market. In turn, a shift in supply for the market as a whole will affect the market price. Entry and exit to and from the market are the driving forces behind a process that, in the long run, pushes the price down to minimum average total costs so that all firms are earning a zero profit.

To understand how short-run profits for a perfectly competitive firm will evaporate in the long run, imagine the following situation. The market is in **long-run equilibrium**, where all firms earn zero economic profits producing the output level where P = MR = MC and P = AC. No firm has the incentive to enter or leave the market. Let's say that the product's demand increases, and with that, the market price goes up. The existing firms in the industry are now facing a higher price than before, so they will increase production to the new output level where P = MR = MC.

This will temporarily make the market price rise above the minimum point on the average cost curve, and therefore, the existing firms in the market will now be earning economic profits. However, these economic profits attract other firms to enter the market. Entry of many new firms causes the market supply curve to shift to the right. As the supply curve shifts to the right, the market price starts decreasing, and with that, economic profits fall for new and existing firms. As long as there are still profits in the market, entry will continue to shift supply to the right. This will stop whenever the market price is driven down to the zero-profit level, where no firm is earning economic profits.

Short-run losses will fade away by reversing this process. Say that the market is in long-run equilibrium. This time, instead, demand decreases, and with that, the market price starts falling. The existing firms in the industry are now facing a lower price than before, and as it will be below the average cost curve, they will now be making economic losses. Some firms will continue producing where the new P = MR = MC, as long as they are able to cover their average variable costs. Some firms will have to shut down immediately as they will not be able to cover their average variable costs, and will then only incur their fixed costs, minimizing their losses. Exit of many firms causes the market supply curve to shift to the left. As the supply curve shifts to the left, the market price starts rising, and economic losses start to be lower. This process ends whenever the market price rises to the zero-profit level, where the existing firms are no longer losing money and are at zero profits again. Thus, while a perfectly competitive firm can earn profits in the short run, in the long run the process of entry will push down prices until they reach the zero-profit level. Conversely, while a perfectly competitive firm may earn losses in the short run, firms will not continually lose money. In the long run, firms making losses are able to escape from their fixed costs, and their exit from the market will push the price back up to the zero-profit level. In the long run, this process of entry and exit will drive the price in perfectly competitive markets to the zero-profit point at the bottom of the AC curve, where marginal cost crosses average cost.

The Long-Run Adjustment and Industry Types

Whenever there are expansions in an industry, costs of production for the existing and new firms could either stay the same, increase, or even decrease. Therefore, we can categorize an industry as being (1) a constant cost industry (as demand increases, the cost of production for firms stays the same), (2) an increasing cost industry (as demand increases, the cost of production for firms increases), or (3) a decreasing cost industry (as demand increases the costs of production for the firms decreases).

For a constant cost industry, whenever there is an increase in market demand and price, then the supply curve shifts to the right with new firms' entry and stops at the point where the new long-run equilibrium intersects at the same market price as before. This is the case of constant returns to scale, which we discussed earlier in the chapter on Production, Costs, and Industry Structure. However, why will costs remain the same? In this type of industry, the supply curve is very elastic. Firms can easily supply any quantity that consumers demand. In addition, there is a perfectly elastic supply of inputs—firms can easily increase their demand for employees, for example, with no increase to wages. Tying in to our Bring it Home discussion, an increased demand for ethanol in recent years has caused the demand for corn to increase. Consequently, many farmers switched from growing wheat to growing corn. Agricultural markets are generally good examples of constant cost industries.

For an increasing cost industry, as the market expands, the old and new firms experience increases in their costs of production, which makes the new zero-profit level intersect at a higher price than before. Here companies may have to deal with limited inputs, such as skilled labor. As the demand for these workers rise, wages rise and this increases the cost of production for all firms. The industry supply curve in this type of industry is more inelastic.

For a decreasing cost industry, as the market expands, the old and new firms experience lower costs of production, which makes the new zero-profit level intersect at a lower price than before. In this case, the industry and all the

firms in it are experiencing falling average total costs. This can be due to an improvement in technology in the entire industry or an increase in the education of employees. High tech industries may be a good example of a decreasing cost market.

Figure 8.8 (a) presents the case of an adjustment process in a constant cost industry. Whenever there are output expansions in this type of industry, the long-run outcome implies more output produced at exactly the same original price. Note that supply was able to increase to meet the increased demand. When we join the before and after long-run equilibriums, the resulting line is the long run supply (LRS) curve in perfectly competitive markets. In this case, it is a flat curve. **Figure 8.8** (b) and **Figure 8.8** (c) present the cases for an increasing cost and decreasing cost industry, respectively. For an increasing cost industry, the LRS is upward sloping, while for a decreasing cost industry, the LRS is downward sloping.

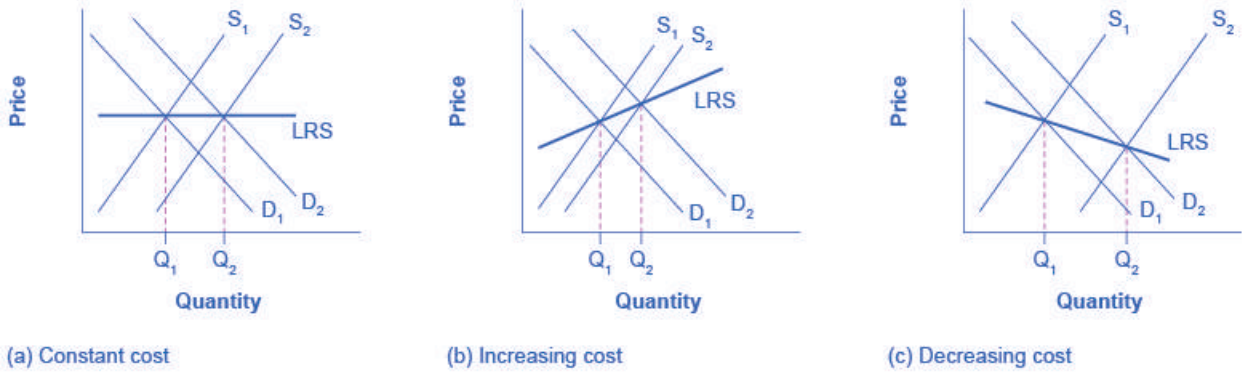


Figure 8.8 Adjustment Process in a Constant-Cost Industry In (a), demand increased and supply met it. Notice that the supply increase is equal to the demand increase. The result is that the equilibrium price stays the same as quantity sold increases. In (b), notice that sellers were not able to increase supply as much as demand. Some inputs were scarce, or wages were rising. The equilibrium price rises. In (c), sellers easily increased supply in response to the demand increase. Here, new technology or economies of scale caused the large increase in supply, resulting in declining equilibrium price.

8.4 | Efficiency in Perfectly Competitive Markets

By the end of this section, you will be able to:

- Apply concepts of productive efficiency and allocative efficiency to perfectly competitive markets
- Compare the model of perfect competition to real-world markets

When profit-maximizing firms in perfectly competitive markets combine with utility-maximizing consumers, something remarkable happens: the resulting quantities of outputs of goods and services demonstrate both productive and allocative efficiency (terms that we first introduced in (**Choice in a World of Scarcity**) .

Productive efficiency means producing without waste, so that the choice is on the production possibility frontier. In the long run in a perfectly competitive market, because of the process of entry and exit, the price in the market is equal to the minimum of the long-run average cost curve. In other words, firms produce and sell goods at the lowest possible average cost.

Allocative efficiency means that among the points on the production possibility frontier, the chosen point is socially preferred—at least in a particular and specific sense. In a perfectly competitive market, price will be equal to the marginal cost of production. Think about the price that one pays for a good as a measure of the social benefit one receives for that good; after all, willingness to pay conveys what the good is worth to a buyer. Then think about the marginal cost of producing the good as representing not just the cost for the firm, but more broadly as the social cost of producing that good. When perfectly competitive firms follow the rule that profits are maximized by producing at the quantity where price is equal to marginal cost, they are thus ensuring that the social benefits they receive from producing a good are in line with the social costs of production.

To explore what economists mean by allocative efficiency, it is useful to walk through an example. Begin by assuming

that the market for wholesale flowers is perfectly competitive, and so P = MC. Now, consider what it would mean if firms in that market produced a lesser quantity of flowers. At a lesser quantity, marginal costs will not yet have increased as much, so that price will exceed marginal cost; that is, P > MC. In that situation, the benefit to society as a whole of producing additional goods, as measured by the willingness of consumers to pay for marginal units of a good, would be higher than the cost of the inputs of labor and physical capital needed to produce the marginal good. In other words, the gains to society as a whole from producing additional marginal units will be greater than the costs.

Conversely, consider what it would mean if, compared to the level of output at the allocatively efficient choice when P = MC, firms produced a greater quantity of flowers. At a greater quantity, marginal costs of production will have increased so that P < MC. In that case, the marginal costs of producing additional flowers is greater than the benefit to society as measured by what people are willing to pay. For society as a whole, since the costs are outstripping the benefits, it will make sense to produce a lower quantity of such goods.

When perfectly competitive firms maximize their profits by producing the quantity where P = MC, they also assure that the benefits to consumers of what they are buying, as measured by the price they are willing to pay, is equal to the costs to society of producing the marginal units, as measured by the marginal costs the firm must pay—and thus that allocative efficiency holds.

We should view the statements that a perfectly competitive market in the long run will feature both productive and allocative efficiency with a degree of skepticism about its truth. Remember, economists are using the concept of "efficiency" in a particular and specific sense, not as a synonym for "desirable in every way." For one thing, consumers' ability to pay reflects the income distribution in a particular society. Thus, a homeless person may have no ability to pay for housing because he or she has insufficient income.

Perfect competition, in the long run, is a hypothetical benchmark. For market structures such as monopoly, monopolistic competition, and oligopoly, which are more frequently observed in the real world than perfect competition, firms will not always produce at the minimum of average cost, nor will they always set price equal to marginal cost. Thus, these other competitive situations will not produce productive and allocative efficiency.

Moreover, real-world markets include many issues that are assumed away in the model of perfect competition, including pollution, inventions of new technology, poverty which may make some people unable to pay for basic necessities of life, government programs like national defense or education, discrimination in labor markets, and buyers and sellers who must deal with imperfect and unclear information. We explore these issues in other chapters. However, the theoretical efficiency of perfect competition does provide a useful benchmark for comparing the issues that arise from these real-world problems.

Bring it Home

A Dime a Dozen

A quick glance at **Table 8.12** reveals the dramatic increase in North Dakota corn production—more than double. Taking into consideration that corn typically yields two to three times as many bushels per acre as wheat, it is obvious there has been a significant increase in bushels of corn. Why the increase in corn acreage? Converging prices.

Year	Corn (millions of acres)	Wheat (millions of acres)
2014	91.6	56.82

Table 8.12 (Source: USDA National Agricultural Statistics Service)

Historically, wheat prices have been higher than corn prices, offsetting wheat's lower yield per acre. However, in recent years wheat and corn prices have been converging. In April 2013, *Agweek* reported the gap was just 71 cents per bushel. As the difference in price narrowed, switching to the production of higher yield per acre of corn simply made good business sense. Erik Younggren, president of the National Association of Wheat Growers said in the *Agweek* article, "I don't think we're going to see mile after mile of waving amber fields [of

wheat] anymore." (Until wheat prices rise, we will probably be seeing field after field of tasseled corn.)

KEY TERMS

break even point level of output where the marginal cost curve intersects the average cost curve at the minimum point of AC; if the price is at this point, the firm is earning zero economic profits

entry the long-run process of firms entering an industry in response to industry profits

exit the long-run process of firms reducing production and shutting down in response to industry losses

long-run equilibrium where all firms earn zero economic profits producing the output level where P = MR = MC and P = AC

marginal revenue the additional revenue gained from selling one more unit

market structure the conditions in an industry, such as number of sellers, how easy or difficult it is for a new firm to enter, and the type of products that are sold

perfect competition each firm faces many competitors that sell identical products

price taker a firm in a perfectly competitive market that must take the prevailing market price as given

shutdown point level of output where the marginal cost curve intersects the average variable cost curve at the minimum point of AVC; if the price is below this point, the firm should shut down immediately

KEY CONCEPTS AND SUMMARY

8.1 Perfect Competition and Why It Matters

A perfectly competitive firm is a price taker, which means that it must accept the equilibrium price at which it sells goods. If a perfectly competitive firm attempts to charge even a tiny amount more than the market price, it will be unable to make any sales. In a perfectly competitive market there are thousands of sellers, easy entry, and identical products. A short-run production period is when firms are producing with some fixed inputs. Long-run equilibrium in a perfectly competitive industry occurs after all firms have entered and exited the industry and seller profits are driven to zero.

Perfect competition means that there are many sellers, there is easy entry and exiting of firms, products are identical from one seller to another, and sellers are price takers.

8.2 How Perfectly Competitive Firms Make Output Decisions

As a perfectly competitive firm produces a greater quantity of output, its total revenue steadily increases at a constant rate determined by the given market price. Profits will be highest (or losses will be smallest) at the quantity of output where total revenues exceed total costs by the greatest amount (or where total revenues fall short of total costs by the smallest amount). Alternatively, profits will be highest where marginal revenue, which is price for a perfectly competitive firm, is equal to marginal cost. If the market price faced by a perfectly competitive firm is above average cost at the profit-maximizing quantity of output, then the firm is making profits. If the market price is below average cost at the profit-maximizing quantity of output, then the firm is making losses.

If the market price is equal to average cost at the profit-maximizing level of output, then the firm is making zero profits. We call the point where the marginal cost curve crosses the average cost curve, at the minimum of the average cost curve, the “zero profit point.” If the market price that a perfectly competitive firm faces is below average variable cost at the profit-maximizing quantity of output, then the firm should shut down operations immediately. If the market price that a perfectly competitive firm faces is above average variable cost, but below average cost, then the firm should continue producing in the short run, but exit in the long run. We call the point where the marginal cost curve crosses the average variable cost curve the shutdown point.

8.3 Entry and Exit Decisions in the Long Run

In the long run, firms will respond to profits through a process of entry, where existing firms expand output and new firms enter the market. Conversely, firms will react to losses in the long run through a process of exit, in which existing firms cease production altogether. Through the process of entry in response to profits and exit in response to losses, the price level in a perfectly competitive market will move toward the zero-profit point, where the marginal cost curve crosses the AC curve at the minimum of the average cost curve.

The long-run supply curve shows the long-run output supplied by firms in three different types of industries: constant cost, increasing cost, and decreasing cost.

8.4 Efficiency in Perfectly Competitive Markets

Long-run equilibrium in perfectly competitive markets meets two important conditions: allocative efficiency and productive efficiency. These two conditions have important implications. First, resources are allocated to their best alternative use. Second, they provide the maximum satisfaction attainable by society.

SELF-CHECK QUESTIONS

1. Firms in a perfectly competitive market are said to be "price takers"—that is, once the market determines an equilibrium price for the product, firms must accept this price. If you sell a product in a perfectly competitive market, but you are not happy with its price, would you raise the price, even by a cent?

2. Would independent trucking fit the characteristics of a perfectly competitive industry?

3. Look at **Table 8.13**. What would happen to the firm's profits if the market price increases to $6 per pack of raspberries?

Quantity	Total Cost	Fixed Cost	Variable Cost	Total Revenue	Profit
0	$62	$62	-	$0	−$62
10	$90	$62	$28	$60	−$30
20	$110	$62	$48	$120	$10
30	$126	$62	$64	$180	$54
40	$144	$62	$82	$240	$96
50	$166	$62	$104	$300	$134
60	$192	$62	$130	$360	$168
70	$224	$62	$162	$420	$196
80	$264	$62	$202	$480	$216
90	$324	$62	$262	$540	$216
100	$404	$62	$342	$600	$196

Table 8.13

4. Suppose that the market price increases to $6, as **Table 8.14** shows. What would happen to the profit-maximizing output level?

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Total Revenue	Marginal Revenue
0	$62	$62	-	-	$0	-
10	$90	$62	$28	$2.80	$60	$6.00
20	$110	$62	$48	$2.00	$120	$6.00
30	$126	$62	$64	$1.60	$180	$6.00
40	$144	$62	$82	$1.80	$240	$6.00
50	$166	$62	$104	$2.20	$300	$6.00
60	$192	$62	$130	$2.60	$360	$6.00
70	$224	$62	$162	$3.20	$420	$6.00
80	$264	$62	$202	$4.00	$480	$6.00
90	$324	$62	$262	**$6.00**	$540	**$6.00**
100	$404	$62	$342	$8.00	$600	$6.00

Table 8.14

5. Explain in words why a profit-maximizing firm will not choose to produce at a quantity where marginal cost exceeds marginal revenue.

6. A firm's marginal cost curve above the average variable cost curve is equal to the firm's individual supply curve. This means that every time a firm receives a price from the market it will be willing to supply the amount of output where the price equals marginal cost. What happens to the firm's individual supply curve if marginal costs increase?

7. If new technology in a perfectly competitive market brings about a substantial reduction in costs of production, how will this affect the market?

8. A market in perfect competition is in long-run equilibrium. What happens to the market if labor unions are able to increase wages for workers?

9. Productive efficiency and allocative efficiency are two concepts achieved in the long run in a perfectly competitive market. These are the two reasons why we call them "perfect." How would you use these two concepts to analyze other market structures and label them "imperfect?"

10. Explain how the profit-maximizing rule of setting P = MC leads a perfectly competitive market to be allocatively efficient.

REVIEW QUESTIONS

11. A single firm in a perfectly competitive market is relatively small compared to the rest of the market. What does this mean? How "small" is "small"?

12. What are the four basic assumptions of perfect competition? Explain in words what they imply for a perfectly competitive firm.

13. What is a "price taker" firm?

14. How does a perfectly competitive firm decide what price to charge?

15. What prevents a perfectly competitive firm from seeking higher profits by increasing the price that it charges?

16. How does a perfectly competitive firm calculate total revenue?

17. Briefly explain the reason for the shape of a marginal revenue curve for a perfectly competitive firm.

18. What two rules does a perfectly competitive firm apply to determine its profit-maximizing quantity of output?

19. How does the average cost curve help to show whether a firm is making profits or losses?

20. What two lines on a cost curve diagram intersect at the zero-profit point?

21. Should a firm shut down immediately if it is making losses?

22. How does the average variable cost curve help a firm know whether it should shut down immediately?

23. What two lines on a cost curve diagram intersect at the shutdown point?

24. Why does entry occur?

25. Why does exit occur?

26. Do entry and exit occur in the short run, the long run, both, or neither?

27. What price will a perfectly competitive firm end up charging in the long run? Why?

28. Will a perfectly competitive market display productive efficiency? Why or why not?

29. Will a perfectly competitive market display allocative efficiency? Why or why not?

CRITICAL THINKING QUESTIONS

30. Finding a life partner is a complicated process that may take many years. It is hard to think of this process as being part of a very complex market, with a demand and a supply for partners. Think about how this market works and some of its characteristics, such as search costs. Would you consider it a perfectly competitive market?

31. Can you name five examples of perfectly competitive markets? Why or why not?

32. Your company operates in a perfectly competitive market. You have been told that advertising can help you increase your sales in the short run. Would you create an aggressive advertising campaign for your product?

33. Since a perfectly competitive firm can sell as much as it wishes at the market price, why can the firm not simply increase its profits by selling an extremely high quantity?

34. Many firms in the United States file for bankruptcy every year, yet they still continue operating. Why would they do this instead of completely shutting down?

35. Why will profits for firms in a perfectly competitive industry tend to vanish in the long run?

36. Why will losses for firms in a perfectly competitive industry tend to vanish in the long run?

37. Assuming that the market for cigarettes is in perfect competition, what does allocative and productive efficiency imply in this case? What does it not imply?

38. In the argument for why perfect competition is allocatively efficient, the price that people are willing to pay represents the gains to society and the marginal cost to the firm represents the costs to society. Can you think of some social costs or issues that are not included in the marginal cost to the firm? Or some social gains that are not included in what people pay for a good?

PROBLEMS

39. The AAA Aquarium Co. sells aquariums for $20 each. Fixed costs of production are $20. The total variable costs are $20 for one aquarium, $25 for two units, $35 for the three units, $50 for four units, and $80 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost, and marginal cost for each output level (one to five units). What is the profit-maximizing quantity of output? On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves.

40. Perfectly competitive firm Doggies Paradise Inc. sells winter coats for dogs. Dog coats sell for $72 each. The fixed costs of production are $100. The total variable costs are $64 for one unit, $84 for two units, $114 for three units, $184 for four units, and $270 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost and marginal cost for each output level (one to five units). On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves. What is the profit maximizing quantity?

41. A computer company produces affordable, easy-to-use home computer systems and has fixed costs of $250. The marginal cost of producing computers is $700 for the first computer, $250 for the second, $300 for the third, $350 for the fourth, $400 for the fifth, $450 for the sixth, and $500 for the seventh.

a. Create a table that shows the company's output, total cost, marginal cost, average cost, variable cost, and average variable cost.
b. At what price is the zero-profit point? At what price is the shutdown point?
c. If the company sells the computers for $500, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.
d. If the firm sells the computers for $300, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.

9 | Monopoly

Figure 9.1 Political Power from a Cotton Monopoly In the mid-nineteenth century, the United States, specifically the Southern states, had a near monopoly in the cotton that they supplied to Great Britain. These states attempted to leverage this economic power into political power—trying to sway Great Britain to formally recognize the Confederate States of America. (Credit: modification of work by "ashleylovespizza"/Flickr Creative Commons)

Bring it Home

The Rest is History

Many of the opening case studies have focused on current events. This one steps into the past to observe how monopoly, or near monopolies, have helped shape history. In spring 1773, the East India Company, a firm that, in its time, was designated "too big to fail," was experiencing financial difficulties. To help shore up the failing firm, the British Parliament authorized the Tea Act. The act continued the tax on teas and made the East India Company the sole legal supplier of tea to the American colonies. By November, the citizens of Boston had had enough. They refused to permit the unloading of tea, citing their main complaint: "No taxation without representation." Several newspapers, including *The Massachusetts Gazette*, warned arriving tea-bearing ships, "We are prepared, and shall not fail to pay them an unwelcome visit by The Mohawks."

Step forward in time to 1860—the eve of the American Civil War—to another near monopoly supplier of historical significance: the U.S. cotton industry. At that time, the Southern states provided the majority of the cotton Britain imported. The South, wanting to secede from the Union, hoped to leverage Britain's high dependency on its cotton into formal diplomatic recognition of the Confederate States of America.

This leads us to this chapter's topic: a firm that controls all (or nearly all) of the supply of a good or service—a monopoly. How do monopoly firms behave in the marketplace? Do they have "power?" Does this power potentially have unintended consequences? We'll return to this case at the end of the chapter to see how the tea and cotton monopolies influenced U.S. history.

Introduction to a Monopoly

In this chapter, you will learn about:

- How Monopolies form: Barriers to Entry
- How a Profit-Maximizing Monopoly Chooses Output and Price

Many believe that top executives at firms are the strongest supporters of market competition, but this belief is far from the truth. Think about it this way: If you very much wanted to win an Olympic gold medal, would you rather be far better than everyone else, or locked in competition with many athletes just as good as you? Similarly, if you would like to attain a very high level of profits, would you rather manage a business with little or no competition, or struggle against many tough competitors who are trying to sell to your customers? By now, you might have read the chapter on **Perfect Competition**. In this chapter, we explore the opposite extreme: monopoly.

If perfect competition is a market where firms have no market power and they simply respond to the market price, monopoly is a market with no competition at all, and firms have a great deal of market power. In the case of **monopoly**, one firm produces all of the output in a market. Since a monopoly faces no significant competition, it can charge any price it wishes, subject to the demand curve. While a monopoly, by definition, refers to a single firm, in practice people often use the term to describe a market in which one firm merely has a very high market share. This tends to be the definition that the U.S. Department of Justice uses.

Even though there are very few true monopolies in existence, we do deal with some of those few every day, often without realizing it: The U.S. Postal Service, your electric, and garbage collection companies are a few examples. Some new drugs are produced by only one pharmaceutical firm—and no close substitutes for that drug may exist.

From the mid-1990s until 2004, the U.S. Department of Justice prosecuted the Microsoft Corporation for including Internet Explorer as the default web browser with its operating system. The Justice Department's argument was that, since Microsoft possessed an extremely high market share in the industry for operating systems, the inclusion of a free web browser constituted unfair competition to other browsers, such as Netscape Navigator. Since nearly everyone was using Windows, including Internet Explorer eliminated the incentive for consumers to explore other browsers and made it impossible for competitors to gain a foothold in the market. In 2013, the Windows system ran on more than 90% of the most commonly sold personal computers. In 2015, a U.S. federal court tossed out antitrust charges that Google had an agreement with mobile device makers to set Google as the default search engine.

This chapter begins by describing how monopolies are protected from competition, including laws that prohibit competition, technological advantages, and certain configurations of demand and supply. It then discusses how a monopoly will choose its profit-maximizing quantity to produce and what price to charge. While a monopoly must be concerned about whether consumers will purchase its products or spend their money on something altogether different, the monopolist need not worry about the actions of other competing firms producing its products. As a result, a monopoly is not a price taker like a perfectly competitive firm, but instead exercises some power to choose its market price.

9.1 | How Monopolies Form: Barriers to Entry

By the end of this section, you will be able to:

- Distinguish between a natural monopoly and a legal monopoly.
- Explain how economies of scale and the control of natural resources led to the necessary formation of legal monopolies
- Analyze the importance of trademarks and patents in promoting innovation
- Identify examples of predatory pricing

Because of the lack of competition, monopolies tend to earn significant economic profits. These profits should attract vigorous competition as we described in **Perfect Competition**, and yet, because of one particular characteristic of monopoly, they do not. **Barriers to entry** are the legal, technological, or market forces that discourage or prevent potential competitors from entering a market. Barriers to entry can range from the simple and easily surmountable,

such as the cost of renting retail space, to the extremely restrictive. For example, there are a finite number of radio frequencies available for broadcasting. Once an entrepreneur or firm has purchased the rights to all of them, no new competitors can enter the market.

In some cases, barriers to entry may lead to monopoly. In other cases, they may limit competition to a few firms. Barriers may block entry even if the firm or firms currently in the market are earning profits. Thus, in markets with significant barriers to entry, it is *not* necessarily true that abnormally high profits will attract new firms, and that this entry of new firms will eventually cause the price to decline so that surviving firms earn only a normal level of profit in the long run.

There are two types of monopoly, based on the types of barriers to entry they exploit. One is **natural monopoly**, where the barriers to entry are something other than legal prohibition. The other is **legal monopoly**, where laws prohibit (or severely limit) competition.

Natural Monopoly

Economies of scale can combine with the size of the market to limit competition. (We introduced this theme in **Production, Cost and Industry Structure**). **Figure 9.2** presents a long-run average cost curve for the airplane manufacturing industry. It shows economies of scale up to an output of 8,000 planes per year and a price of P_0, then constant returns to scale from 8,000 to 20,000 planes per year, and diseconomies of scale at a quantity of production greater than 20,000 planes per year.

Now consider the market demand curve in the diagram, which intersects the long-run average cost (LRAC) curve at an output level of 5,000 planes per year and at a price P_1, which is higher than P_0. In this situation, the market has room for only one producer. If a second firm attempts to enter the market at a smaller size, say by producing a quantity of 4,000 planes, then its average costs will be higher than those of the existing firm, and it will be unable to compete. If the second firm attempts to enter the market at a larger size, like 8,000 planes per year, then it could produce at a lower average cost—but it could not sell all 8,000 planes that it produced because of insufficient demand in the market.

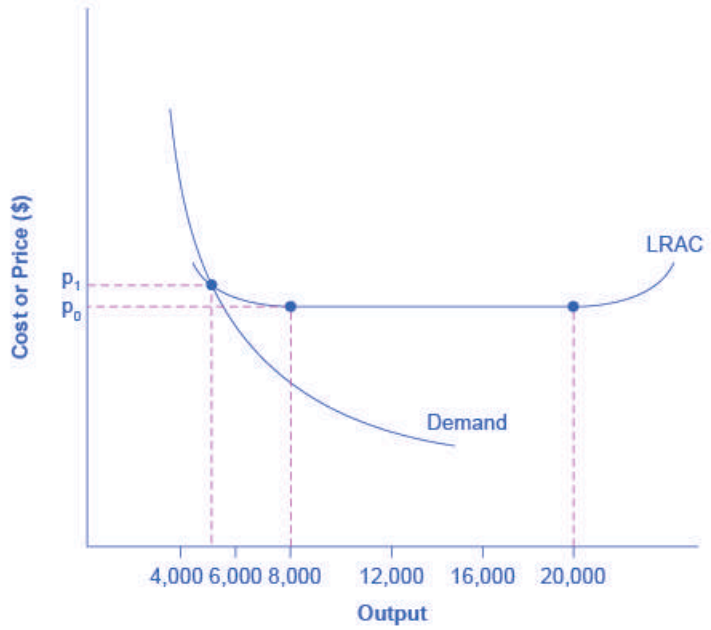


Figure 9.2 Economies of Scale and Natural Monopoly In this market, the demand curve intersects the long-run average cost (LRAC) curve at its downward-sloping part. A natural monopoly occurs when the quantity demanded is less than the minimum quantity it takes to be at the bottom of the long-run average cost curve.

Economists call this situation, when economies of scale are large relative to the quantity demanded in the market, a natural monopoly. Natural monopolies often arise in industries where the marginal cost of adding an additional customer is very low, once the fixed costs of the overall system are in place. This results in situations where there are substantial economies of scale. For example, once a water company lays the main water pipes through a neighborhood, the marginal cost of providing water service to another home is fairly low. Once the electric company

installs lines in a new subdivision, the marginal cost of providing additional electrical service to one more home is minimal. It would be costly and duplicative for a second water company to enter the market and invest in a whole second set of main water pipes, or for a second electricity company to enter the market and invest in a whole new set of electrical wires. These industries offer an example where, because of economies of scale, one producer can serve the entire market more efficiently than a number of smaller producers that would need to make duplicate physical capital investments.

A natural monopoly can also arise in smaller local markets for products that are difficult to transport. For example, cement production exhibits economies of scale, and the quantity of cement demanded in a local area may not be much larger than what a single plant can produce. Moreover, the costs of transporting cement over land are high, and so a cement plant in an area without access to water transportation may be a natural monopoly.

Control of a Physical Resource

Another type of natural monopoly occurs when a company has control of a scarce physical resource. In the U.S. economy, one historical example of this pattern occurred when ALCOA—the Aluminum Company of America—controlled most of the supply of bauxite, a key mineral used in making aluminum. Back in the 1930s, when ALCOA controlled most of the bauxite, other firms were simply unable to produce enough aluminum to compete.

As another example, the majority of global diamond production is controlled by DeBeers, a multi-national company that has mining and production operations in South Africa, Botswana, Namibia, and Canada. It also has exploration activities on four continents, while directing a worldwide distribution network of rough cut diamonds. Although in recent years they have experienced growing competition, their impact on the rough diamond market is still considerable.

Legal Monopoly

For some products, the government erects barriers to entry by prohibiting or limiting competition. Under U.S. law, no organization but the U.S. Postal Service is legally allowed to deliver first-class mail. Many states or cities have laws or regulations that allow households a choice of only one electric company, one water company, and one company to pick up the garbage. Most legal monopolies are utilities—products necessary for everyday life—that are socially beneficial. As a consequence, the government allows producers to become regulated monopolies, to insure that customers have access to an appropriate amount of these products or services. Additionally, legal monopolies are often subject to economies of scale, so it makes sense to allow only one provider.

Promoting Innovation

Innovation takes time and resources to achieve. Suppose a company invests in research and development and finds the cure for the common cold. In this world of near ubiquitous information, other companies could take the formula, produce the drug, and because they did not incur the costs of research and development (R&D), undercut the price of the company that discovered the drug. Given this possibility, many firms would choose not to invest in research and development, and as a result, the world would have less innovation. To prevent this from happening, the Constitution of the United States specifies in Article I, Section 8: “The Congress shall have Power . . . to Promote the Progress of Science and Useful Arts, by securing for limited Times to Authors and Inventors the Exclusive Right to their Writings and Discoveries.” Congress used this power to create the U.S. Patent and Trademark Office, as well as the U.S. Copyright Office. A **patent** gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time. In the United States, exclusive patent rights last for 20 years. The idea is to provide limited monopoly power so that innovative firms can recoup their investment in R&D, but then to allow other firms to produce the product more cheaply once the patent expires.

A **trademark** is an identifying symbol or name for a particular good, like Chiquita bananas, Chevrolet cars, or the Nike “swoosh” that appears on shoes and athletic gear. Roughly 1.9 million trademarks are registered with the U.S. government. A firm can renew a trademark repeatedly, as long as it remains in active use.

A **copyright**, according to the U.S. Copyright Office, “is a form of protection provided by the laws of the United States for ‘original works of authorship’ including literary, dramatic, musical, architectural, cartographic, choreographic, pantomimic, pictorial, graphic, sculptural, and audiovisual creations.” No one can reproduce, display, or perform a copyrighted work without the author's permission. Copyright protection ordinarily lasts for the life of the author plus 70 years.

Roughly speaking, patent law covers inventions and copyright protects books, songs, and art. However, in certain

areas, like the invention of new software, it has been unclear whether patent or copyright protection should apply. There is also a body of law known as **trade secrets**. Even if a company does not have a patent on an invention, competing firms are not allowed to steal their secrets. One famous trade secret is the formula for Coca-Cola, which is not protected under copyright or patent law, but is simply kept secret by the company.

Taken together, we call this combination of patents, trademarks, copyrights, and trade secret law **intellectual property**, because it implies ownership over an idea, concept, or image, not a physical piece of property like a house or a car. Countries around the world have enacted laws to protect intellectual property, although the time periods and exact provisions of such laws vary across countries. There are ongoing negotiations, both through the World Intellectual Property Organization (WIPO) and through international treaties, to bring greater harmony to the intellectual property laws of different countries to determine the extent to which those in other countries will respect patents and copyrights of those in other countries.

Government limitations on competition used to be more common in the United States. For most of the twentieth century, only one phone company—AT&T—was legally allowed to provide local and long distance service. From the 1930s to the 1970s, one set of federal regulations limited which destinations airlines could choose to fly to and what fares they could charge. Another set of regulations limited the interest rates that banks could pay to depositors; yet another specified how much trucking firms could charge customers.

What products we consider utilities depends, in part, on the available technology. Fifty years ago, telephone companies provided local and long distance service over wires. It did not make much sense to have many companies building multiple wiring systems across towns and the entire country. AT&T lost its monopoly on long distance service when the technology for providing phone service changed from wires to microwave and satellite transmission, so that multiple firms could use the same transmission mechanism. The same thing happened to local service, especially in recent years, with the growth in cellular phone systems.

The combination of improvements in production technologies and a general sense that the markets could provide services adequately led to a wave of **deregulation**, starting in the late 1970s and continuing into the 1990s. This wave eliminated or reduced government restrictions on the firms that could enter, the prices that they could charge, and the quantities that many industries could produce, including telecommunications, airlines, trucking, banking, and electricity.

Around the world, from Europe to Latin America to Africa and Asia, many governments continue to control and limit competition in what those governments perceive to be key industries, including airlines, banks, steel companies, oil companies, and telephone companies.

Link It Up

Vist this website (http://openstaxcollege.org/l/patents) for examples of some pretty bizarre patents.

Intimidating Potential Competition

Businesses have developed a number of schemes for creating barriers to entry by deterring potential competitors from entering the market. One method is known as **predatory pricing**, in which a firm uses the threat of sharp price cuts to discourage competition. Predatory pricing is a violation of U.S. antitrust law, but it is difficult to prove.

Consider a large airline that provides most of the flights between two particular cities. A new, small start-up airline decides to offer service between these two cities. The large airline immediately slashes prices on this route to the bone, so that the new entrant cannot make any money. After the new entrant has gone out of business, the incumbent

firm can raise prices again.

After the company repeats this pattern once or twice, potential new entrants may decide that it is not wise to try to compete. Small airlines often accuse larger airlines of predatory pricing: in the early 2000s, for example, ValuJet accused Delta of predatory pricing, Frontier accused United, and Reno Air accused Northwest. In 2015, the Justice Department ruled against American Express and Mastercard for imposing restrictions on retailers that encouraged customers to use lower swipe fees on credit transactions.

In some cases, large advertising budgets can also act as a way of discouraging the competition. If the only way to launch a successful new national cola drink is to spend more than the promotional budgets of Coca-Cola and Pepsi Cola, not too many companies will try. A firmly established brand name can be difficult to dislodge.

Summing Up Barriers to Entry

Table 9.1 lists the barriers to entry that we have discussed. This list is not exhaustive, since firms have proved to be highly creative in inventing business practices that discourage competition. When barriers to entry exist, perfect competition is no longer a reasonable description of how an industry works. When barriers to entry are high enough, monopoly can result.

Barrier to Entry	Government Role?	Example
Natural monopoly	Government often responds with regulation (or ownership)	Water and electric companies
Control of a physical resource	No	DeBeers for diamonds
Legal monopoly	Yes	Post office, past regulation of airlines and trucking
Patent, trademark, and copyright	Yes, through protection of intellectual property	New drugs or software
Intimidating potential competitors	Somewhat	Predatory pricing; well-known brand names

Table 9.1 Barriers to Entry

9.2 | How a Profit-Maximizing Monopoly Chooses Output and Price

By the end of this section, you will be able to:

- Explain the perceived demand curve for a perfect competitor and a monopoly
- Analyze a demand curve for a monopoly and determine the output that maximizes profit and revenue
- Calculate marginal revenue and marginal cost
- Explain allocative efficiency as it pertains to the efficiency of a monopoly

Consider a monopoly firm, comfortably surrounded by barriers to entry so that it need not fear competition from other producers. How will this monopoly choose its profit-maximizing quantity of output, and what price will it charge? Profits for the monopolist, like any firm, will be equal to total revenues minus total costs. We can analyze the pattern of costs for the monopoly within the same framework as the costs of a perfectly competitive firm—that is, by using total cost, fixed cost, variable cost, marginal cost, average cost, and average variable cost. However, because a monopoly faces no competition, its situation and its decision process will differ from that of a perfectly competitive firm. (The Clear It Up feature discusses how hard it is sometimes to define “market” in a monopoly situation.)

Demand Curves Perceived by a Perfectly Competitive Firm and by a Monopoly

A perfectly competitive firm acts as a price taker, so we calculate total revenue taking the given market price and multiplying it by the quantity of output that the firm chooses. The demand curve *as it is perceived by a perfectly competitive firm* appears in **Figure 9.3** (a). The flat perceived demand curve means that, from the viewpoint of the perfectly competitive firm, it could sell either a relatively low quantity like Ql or a relatively high quantity like Qh at the market price P.

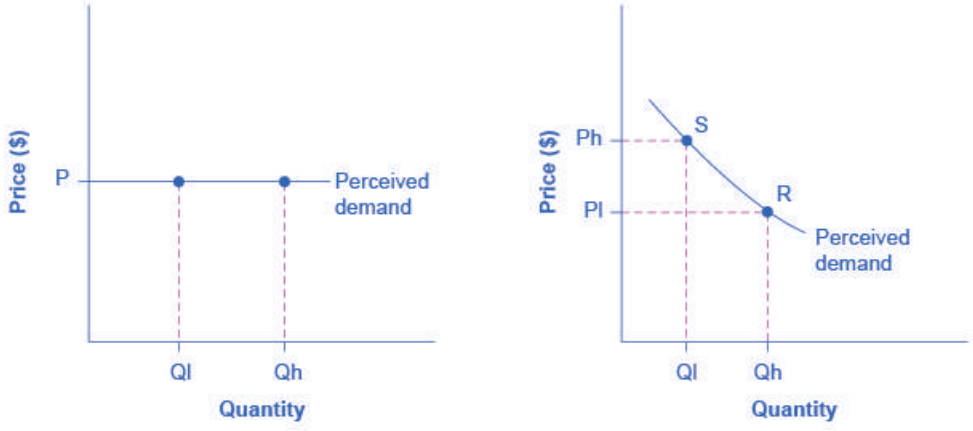


Figure 9.3 The Perceived Demand Curve for a Perfect Competitor and a Monopolist (a) A perfectly competitive firm perceives the demand curve that it faces to be flat. The flat shape means that the firm can sell either a low quantity (Ql) or a high quantity (Qh) at exactly the same price (P). (b) A monopolist perceives the demand curve that it faces to be the same as the market demand curve, which for most goods is downward-sloping. Thus, if the monopolist chooses a high level of output (Qh), it can charge only a relatively low price (Pl). Conversely, if the monopolist chooses a low level of output (Ql), it can then charge a higher price (Ph). The challenge for the monopolist is to choose the combination of price and quantity that maximizes profits.

Clear It Up

What defines the market?

A monopoly is a firm that sells all or nearly all of the goods and services in a given market. However, what defines the "market"?

In a famous 1947 case, the federal government accused the DuPont company of having a monopoly in the cellophane market, pointing out that DuPont produced 75% of the cellophane in the United States. DuPont countered that even though it had a 75% market share in cellophane, it had less than a 20% share of the "flexible packaging materials," which includes all other moisture-proof papers, films, and foils. In 1956, after years of legal appeals, the U.S. Supreme Court held that the broader market definition was more appropriate, and it dismissed the case against DuPont.

Questions over how to define the market continue today. True, Microsoft in the 1990s had a dominant share of the software for computer operating systems, but in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. The Greyhound bus company may have a near-monopoly on the market for intercity bus transportation, but it is only a small share of the market for intercity transportation if that market includes private cars, airplanes, and railroad service. DeBeers has a monopoly in diamonds, but it is a much smaller share of the total market for precious gemstones and an even smaller share of the total market for jewelry. A small town in the country may have only one gas station: is this gas station a "monopoly," or does it compete with gas stations that might be

five, 10, or 50 miles away?

In general, if a firm produces a product without close substitutes, then we can consider the firm a monopoly producer in a single market. However, if buyers have a range of similar—even if not identical—options available from other firms, then the firm is not a monopoly. Still, arguments over whether substitutes are close or not close can be controversial.

While a monopolist can charge *any* price for its product, nonetheless the demand for the firm's product constrains the price. No monopolist, even one that is thoroughly protected by high barriers to entry, can require consumers to purchase its product. Because the monopolist is the only firm in the market, its demand curve is the same as the market demand curve, which is, unlike that for a perfectly competitive firm, downward-sloping.

Figure 9.3 illustrates this situation. The monopolist can either choose a point like R with a low price (Pl) and high quantity (Qh), or a point like S with a high price (Ph) and a low quantity (Ql), or some intermediate point. Setting the price too high will result in a low quantity sold, and will not bring in much revenue. Conversely, setting the price too low may result in a high quantity sold, but because of the low price, it will not bring in much revenue either. The challenge for the monopolist is to strike a profit-maximizing balance between the price it charges and the quantity that it sells. However, why isn't the perfectly competitive firm's demand curve also the market demand curve? See the following Clear It Up feature for the answer to this question.

Clear It Up

What is the difference between perceived demand and market demand?

The demand curve as perceived by a perfectly competitive firm is not the overall market demand curve for that product. However, the firm's demand curve as perceived by a monopoly is the same as the market demand curve. The reason for the difference is that each perfectly competitive firm perceives the demand for its products in a market that includes many other firms. In effect, the demand curve perceived by a perfectly competitive firm is a tiny slice of the entire market demand curve. In contrast, a monopoly perceives demand for its product in a market where the monopoly is the only producer.

Total Cost and Total Revenue for a Monopolist

We can illustrate profits for a monopolist with a graph of total revenues and total costs, with the example of the hypothetical HealthPill firm in **Figure 9.4**. The total cost curve has its typical shape that we learned about in **Production, Costs and Industry Structure**, and that we used in **Perfect Competition**; that is, total costs rise and the curve grows steeper as output increases, as the final column of **Table 9.2** shows.

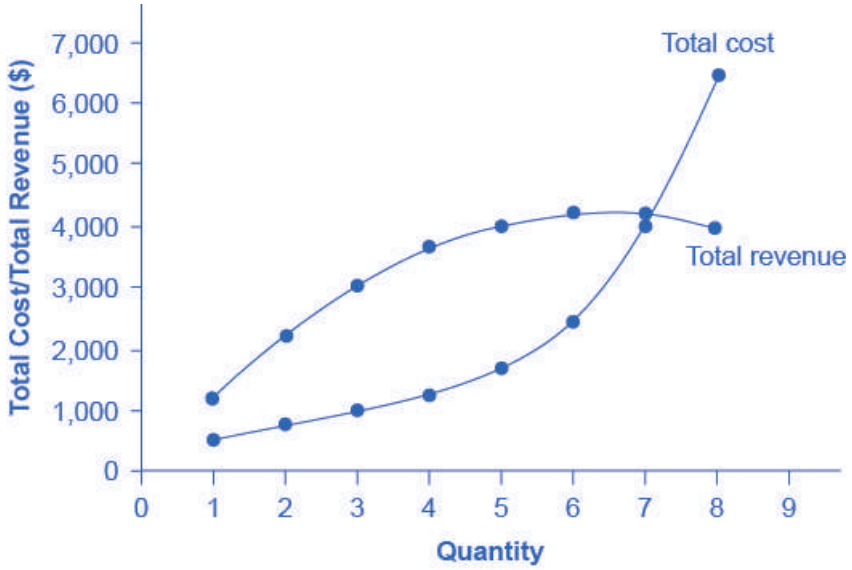


Figure 9.4 Total Revenue and Total Cost for the HealthPill Monopoly Total revenue for the monopoly firm called HealthPill first rises, then falls. Low levels of output bring in relatively little total revenue, because the quantity is low. High levels of output bring in relatively less revenue, because the high quantity pushes down the market price. The total cost curve is upward-sloping. Profits will be highest at the quantity of output where total revenue is most above total cost. The profit-maximizing level of output is not the same as the revenue-maximizing level of output, which should make sense, because profits take costs into account and revenues do not.

Quantity Q	Price P	Total Revenue TR	Total Cost TC
1	1,200	1,200	500
2	1,100	2,200	750
3	1,000	3,000	1,000
4	900	3,600	1,250
5	800	4,000	1,650
6	700	4,200	2,500
7	600	4,200	4,000
8	500	4,000	6,400

Table 9.2 Total Costs and Total Revenues of HealthPill

Total revenue, though, is different. Since a monopolist faces a downward sloping demand curve, the only way it can sell more output is by reducing its price. Selling more output raises revenue, but lowering price reduces it. Thus, the shape of total revenue isn't clear. Let's explore this using the data in **Table 9.2**, which shows quantities along the demand curve and the price at each quantity demanded, and then calculates total revenue by multiplying price times quantity at each level of output. (In this example, we give the output as 1, 2, 3, 4, and so on, for the sake of simplicity. If you prefer a dash of greater realism, you can imagine that the pharmaceutical company measures the pharmaceutical company measures these output levels and the corresponding prices per 1,000 or 10,000 pills.) As the figure illustrates, total revenue for a monopolist has the shape of a hill, first rising, next flattening out, and then falling. In this example, total revenue is highest at a quantity of 6 or 7.

However, the monopolist is not seeking to maximize revenue, but instead to earn the highest possible profit. In the

HealthPill example in **Figure 9.4**, the highest profit will occur at the quantity where total revenue is the farthest above total cost. This looks to be somewhere in the middle of the graph, but where exactly? It is easier to see the profit maximizing level of output by using the marginal approach, to which we turn next.

Marginal Revenue and Marginal Cost for a Monopolist

In the real world, a monopolist often does not have enough information to analyze its entire total revenues or total costs curves. After all, the firm does not know exactly what would happen if it were to alter production dramatically. However, a monopolist often has fairly reliable information about how changing output by small or moderate amounts will affect its marginal revenues and marginal costs, because it has had experience with such changes over time and because modest changes are easier to extrapolate from current experience. A monopolist can use information on marginal revenue and marginal cost to seek out the profit-maximizing combination of quantity and price.

Table 9.3 expands **Table 9.2** using the figures on total costs and total revenues from the HealthPill example to calculate marginal revenue and marginal cost. This monopoly faces typical upward-sloping marginal cost and downward sloping marginal revenue curves, as **Figure 9.5** shows.

Notice that marginal revenue is zero at a quantity of 7, and turns negative at quantities higher than 7. It may seem counterintuitive that marginal revenue could ever be zero or negative: after all, doesn't an increase in quantity sold not always mean more revenue? For a perfect competitor, each additional unit sold brought a positive marginal revenue, because marginal revenue was equal to the given market price. However, a monopolist can sell a larger quantity and see a decline in total revenue. When a monopolist increases sales by one unit, it gains some marginal revenue from selling that extra unit, but also loses some marginal revenue because it must now sell every other unit at a lower price. As the quantity sold becomes higher, at some point the drop in price is proportionally more than the increase in greater quantity of sales, causing a situation where more sales bring in less revenue. In other words, marginal revenue is negative.

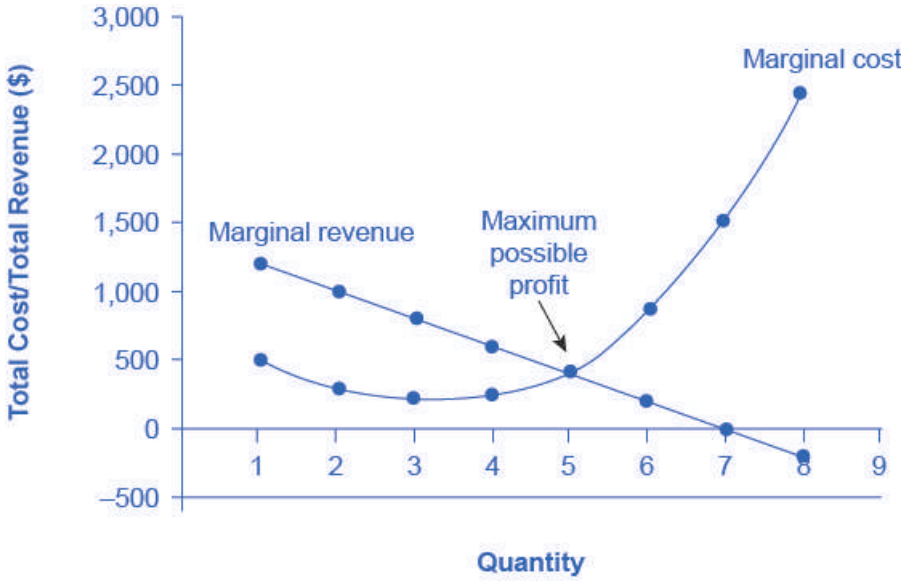


Figure 9.5 Marginal Revenue and Marginal Cost for the HealthPill Monopoly For a monopoly like HealthPill, marginal revenue decreases as it sells additional units of output. The marginal cost curve is upward-sloping. The profit-maximizing choice for the monopoly will be to produce at the quantity where marginal revenue is equal to marginal cost: that is, MR = MC. If the monopoly produces a lower quantity, then MR > MC at those levels of output, and the firm can make higher profits by expanding output. If the firm produces at a greater quantity, then MC > MR, and the firm can make higher profits by reducing its quantity of output.

Quantity Q	Total Revenue TR	Marginal Revenue MR	Total Cost TC	Marginal Cost MC
1	1,200	1,200	500	500
2	2,200	1,000	775	275
3	3,000	800	1,000	225
4	3,600	600	1,250	250
5	4,000	400	1,650	400
6	4,200	200	2,500	850
7	4,200	0	4,000	1,500
8	4,000	–200	6,400	2,400

Table 9.3 Costs and Revenues of HealthPill

A monopolist can determine its profit-maximizing price and quantity by analyzing the marginal revenue and marginal costs of producing an extra unit. If the marginal revenue exceeds the marginal cost, then the firm should produce the extra unit.

For example, at an output of 4 in **Figure 9.5**, marginal revenue is 600 and marginal cost is 250, so producing this unit will clearly add to overall profits. At an output of 5, marginal revenue is 400 and marginal cost is 400, so producing this unit still means overall profits are unchanged. However, expanding output from 5 to 6 would involve a marginal revenue of 200 and a marginal cost of 850, so that sixth unit would actually reduce profits. Thus, the monopoly can tell from the marginal revenue and marginal cost that of the choices in the table, the profit-maximizing level of output is 5.

The monopoly could seek out the profit-maximizing level of output by increasing quantity by a small amount, calculating marginal revenue and marginal cost, and then either increasing output as long as marginal revenue exceeds marginal cost or reducing output if marginal cost exceeds marginal revenue. This process works without any need to calculate total revenue and total cost. Thus, a profit-maximizing monopoly should follow the rule of producing up to the quantity where marginal revenue is equal to marginal cost—that is, MR = MC. This quantity is easy to identify graphically, where MR and MC intersect.

Work It Out

Maximizing Profits

If you find it counterintuitive that producing where marginal revenue equals marginal cost will maximize profits, working through the numbers will help.

Step 1. Remember, we define marginal cost as the change in total cost from producing a small amount of additional output.

$$\text{MC} = \frac{\text{change in total cost}}{\text{change in quantity produced}}$$

Step 2. Note that in **Table 9.3**, as output increases from 1 to 2 units, total cost increases from $1500 to $1800. As a result, the marginal cost of the second unit will be:

$$\begin{aligned} \text{MC} &= \frac{\$775 - \$500}{1} \\ &= \$275 \end{aligned}$$

Step 3. Remember that, similarly, marginal revenue is the change in total revenue from selling a small amount of additional output.

$$\text{MR} = \frac{\text{change in total revenue}}{\text{change in quantity sold}}$$

Step 4. Note that in **Table 9.3**, as output increases from 1 to 2 units, total revenue increases from \$1200 to \$2200. As a result, the marginal revenue of the second unit will be:

$$\text{MR} = \frac{\$2200 - \$1200}{1} = \$1000$$

Quantity Q	Marginal Revenue MR	Marginal Cost MC	Marginal Profit MP	Total Profit P
1	1,200	500	700	700
2	1,000	275	725	1,425
3	800	225	575	2,000
4	600	250	350	2,350
5	400	400	0	2,350
6	200	850	−650	1,700
7	0	1,500	−1,500	200
8	−200	2,400	−2,600	−2,400

Table 9.4 Marginal Revenue, Marginal Cost, Marginal and Total Profit

Table 9.4 repeats the marginal cost and marginal revenue data from **Table 9.3**, and adds two more columns: **Marginal profit** is the profitability of each additional unit sold. We define it as marginal revenue minus marginal cost. Finally, total profit is the sum of marginal profits. As long as marginal profit is positive, producing more output will increase total profits. When marginal profit turns negative, producing more output will decrease total profits. Total profit is maximized where marginal revenue equals marginal cost. In this example, maximum profit occurs at 5 units of output.

A perfectly competitive firm will also find its profit-maximizing level of output where MR = MC. The key difference with a perfectly competitive firm is that in the case of perfect competition, marginal revenue is equal to price (MR = P), while for a monopolist, marginal revenue is not equal to the price, because changes in quantity of output affect the price.

Illustrating Monopoly Profits

It is straightforward to calculate profits of given numbers for total revenue and total cost. However, the size of monopoly profits can also be illustrated graphically with **Figure 9.6**, which takes the marginal cost and marginal revenue curves from the previous exhibit and adds an average cost curve and the monopolist's perceived demand curve. **Table 9.5** shows the data for these curves.

Quantity Q	Demand P	Marginal Revenue MR	Marginal Cost MC	Average Cost AC
1	1,200	1,200	500	500
2	1,100	1,000	275	388
3	1,000	800	225	333
4	900	600	250	313
5	800	400	400	330
6	700	200	850	417
7	600	0	1,500	571
8	500	–200	2,400	800

Table 9.5

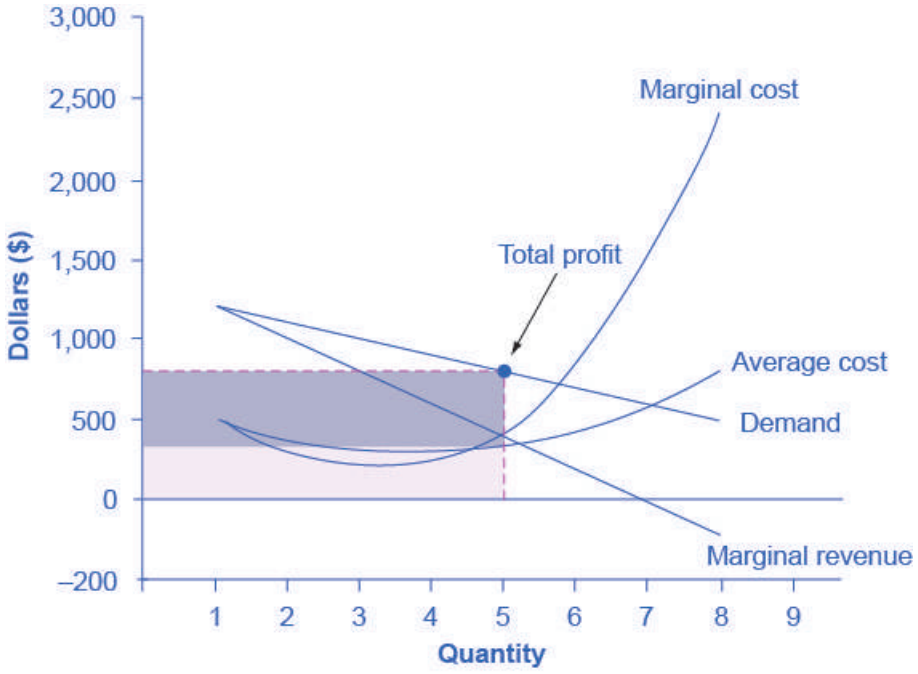


Figure 9.6 Illustrating Profits at the HealthPill Monopoly This figure begins with the same marginal revenue and marginal cost curves from the HealthPill monopoly from **Figure 9.5**. It then adds an average cost curve and the demand curve that the monopolist faces. The HealthPill firm first chooses the quantity where MR = MC. In this example, the quantity is 5. The monopolist then decides what price to charge by looking at the demand curve it faces. The large box, with quantity on the horizontal axis and demand (which shows the price) on the vertical axis, shows total revenue for the firm. The lighter-shaded box, which is quantity on the horizontal axis and average cost of production on the vertical axis shows the firm's total costs. The large total revenue box minus the smaller total cost box leaves the darkly shaded box that shows total profits. Since the price charged is above average cost, the firm is earning positive profits.

Figure 9.7 illustrates the three-step process where a monopolist: selects the profit-maximizing quantity to produce; decides what price to charge; determines total revenue, total cost, and profit.

Step 1: The Monopolist Determines Its Profit-Maximizing Level of Output

The firm can use the points on the demand curve D to calculate total revenue, and then, based on total revenue,

calculate its marginal revenue curve. The profit-maximizing quantity will occur where MR = MC—or at the last possible point before marginal costs start exceeding marginal revenue. On **Figure 9.6**, MR = MC occurs at an output of 5.

Step 2: The Monopolist Decides What Price to Charge

The monopolist will charge what the market is willing to pay. A dotted line drawn straight up from the profit-maximizing quantity to the demand curve shows the profit-maximizing price which, in **Figure 9.6**, is $800. This price is above the average cost curve, which shows that the firm is earning profits.

Step 3: Calculate Total Revenue, Total Cost, and Profit

Total revenue is the overall shaded box, where the width of the box is the quantity sold and the height is the price. In **Figure 9.6**, this is 5 x $800 = $4000. In **Figure 9.6**, the bottom part of the shaded box, which is shaded more lightly, shows total costs; that is, quantity on the horizontal axis multiplied by average cost on the vertical axis or 5 x $330 = $1650. The larger box of total revenues minus the smaller box of total costs will equal profits, which the darkly shaded box shows. Using the numbers gives $4000 - $1650 = $2350. In a perfectly competitive market, the forces of entry would erode this profit in the long run. However, a monopolist is protected by barriers to entry. In fact, one obvious sign of a possible monopoly is when a firm earns profits year after year, while doing more or less the same thing, without ever seeing increased competition eroding those profits.

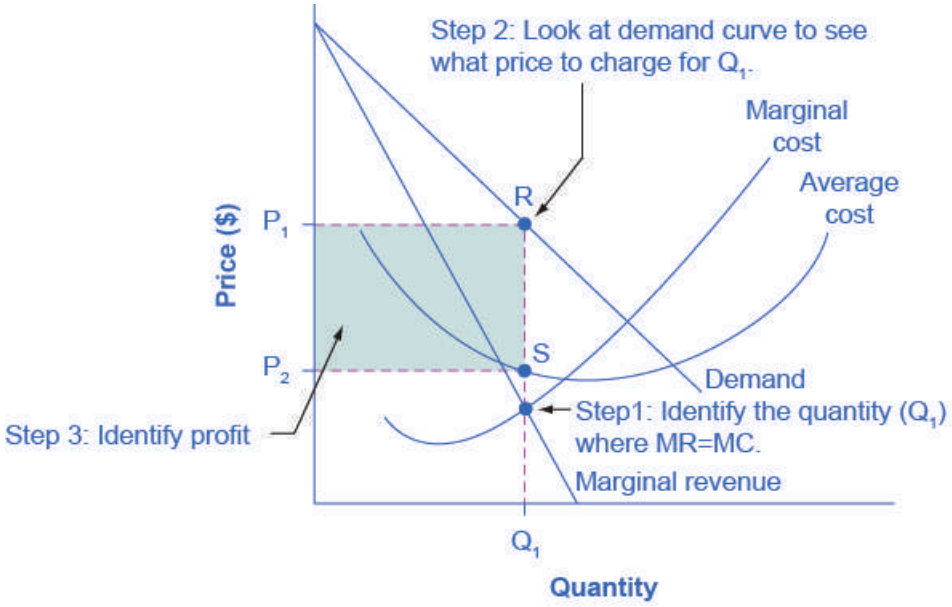


Figure 9.7 How a Profit-Maximizing Monopoly Decides Price In Step 1, the monopoly chooses the profit-maximizing level of output Q_1, by choosing the quantity where MR = MC. In Step 2, the monopoly decides how much to charge for output level Q_1 by drawing a line straight up from Q_1 to point R on its perceived demand curve. Thus, the monopoly will charge a price (P_1). In Step 3, the monopoly identifies its profit. Total revenue will be Q_1 multiplied by P_1. Total cost will be Q_1 multiplied by the average cost of producing Q_1, which point S shows on the average cost curve to be P_2. Profits will be the total revenue rectangle minus the total cost rectangle, which the shaded zone in the figure shows.

Clear It Up

Why is a monopolist's marginal revenue always less than the price?

The marginal revenue curve for a monopolist always lies beneath the market demand curve. To understand why, think about increasing the quantity along the demand curve by one unit, so that you take one step down

the demand curve to a slightly higher quantity but a slightly lower price. A demand curve is not sequential: It is not that first we sell Q_1 at a higher price, and then we sell Q_2 at a lower price. Rather, a demand curve is conditional: If we charge the higher price, we would sell Q_1. If, instead, we charge a lower price (on all the units that we sell), we would sell Q_2.

When we think about increasing the quantity sold by one unit, marginal revenue is affected in two ways. First, we sell one additional unit at the new market price. Second, all the previous units, which we sold at the higher price, now sell for less. Because of the lower price on all units sold, the marginal revenue of selling a unit is less than the price of that unit—and the marginal revenue curve is below the demand curve. *Tip*: For a straight-line demand curve, MR and demand have the same vertical intercept. As output increases, marginal revenue decreases twice as fast as demand, so that the horizontal intercept of MR is halfway to the horizontal intercept of demand. You can see this in the **Figure 9.8**.

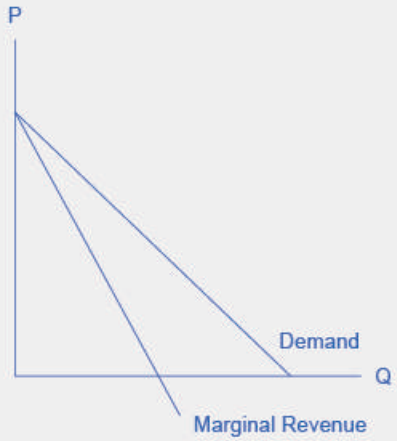


Figure 9.8 The Monopolist's Marginal Revenue Curve versus Demand Curve Because the market demand curve is conditional, the marginal revenue curve for a monopolist lies beneath the demand curve.

The Inefficiency of Monopoly

Most people criticize monopolies because they charge too high a price, but what economists object to is that monopolies do not supply enough output to be allocatively efficient. To understand why a monopoly is inefficient, it is useful to compare it with the benchmark model of perfect competition.

Allocative efficiency is an economic concept regarding efficiency at the social or societal level. It refers to producing the optimal quantity of some output, the quantity where the marginal benefit to society of one more unit just equals the marginal cost. The rule of profit maximization in a world of perfect competition was for each firm to produce the quantity of output where P = MC, where the price (P) is a measure of how much buyers value the good and the marginal cost (MC) is a measure of what marginal units cost society to produce. Following this rule assures allocative efficiency. If P > MC, then the marginal benefit to society (as measured by P) is greater than the marginal cost to society of producing additional units, and a greater quantity should be produced. However, in the case of monopoly, price is always greater than marginal cost at the profit-maximizing level of output, as you can see by looking back at **Figure 9.6**. Thus, consumers will suffer from a monopoly because it will sell a lower quantity in the market, at a higher price, than would have been the case in a perfectly competitive market.

The problem of inefficiency for monopolies often runs even deeper than these issues, and also involves incentives for efficiency over longer periods of time. There are counterbalancing incentives here. On one side, firms may strive for new inventions and new intellectual property because they want to become monopolies and earn high profits—at least for a few years until the competition catches up. In this way, monopolies may come to exist because of competitive pressures on firms. However, once a barrier to entry is in place, a monopoly that does not need to fear competition can just produce the same old products in the same old way—while still ringing up a healthy rate of profit. John Hicks, who won the Nobel Prize for economics in 1972, wrote in 1935: "The best of all monopoly profits is a quiet life." He did not mean the comment in a complimentary way. He meant that monopolies may bank their profits and slack off on trying to please their customers.

When AT&T provided all of the local and long-distance phone service in the United States, along with manufacturing most of the phone equipment, the payment plans and types of phones did not change much. The old joke was that you could have any color phone you wanted, as long as it was black. However, in 1982, government litigation split up AT&T into a number of local phone companies, a long-distance phone company, and a phone equipment manufacturer. An explosion of innovation followed. Services like call waiting, caller ID, three-way calling, voice mail through the phone company, mobile phones, and wireless connections to the internet all became available. Companies offered a wide range of payment plans, as well. It was no longer true that all phones were black. Instead, phones came in a wide variety of shapes and colors. The end of the telephone monopoly brought lower prices, a greater quantity of services, and also a wave of innovation aimed at attracting and pleasing customers.

Bring it Home

The Rest is History

In the opening case, we presented the East India Company and the Confederate States as a monopoly or near monopoly provider of a good. Nearly every American schoolchild knows the result of the “unwelcome visit” the “Mohawks” bestowed upon Boston Harbor’s tea-bearing ships—the Boston Tea Party. Regarding the cotton industry, we also know Great Britain remained neutral during the Civil War, taking neither side during the conflict.

Did the monopoly nature of these business have unintended and historical consequences? Might the American Revolution have been deterred, if the East India Company had sailed the tea-bearing ships back to England? Might the southern states have made different decisions had they not been so confident “King Cotton” would force diplomatic recognition of the Confederate States of America? Of course, it is not possible to definitively answer these questions. We cannot roll back the clock and try a different scenario. We can, however, consider the monopoly nature of these businesses and the roles they played and hypothesize about what might have occurred under different circumstances.

Perhaps if there had been legal free tea trade, the colonists would have seen things differently. There was smuggled Dutch tea in the colonial market. If the colonists had been able to freely purchase Dutch tea, they would have paid lower prices and avoided the tax.

What about the cotton monopoly? With one in five jobs in Great Britain depending on Southern cotton and the Confederate States as nearly the sole provider of that cotton, why did Great Britain remain neutral during the Civil War? At the beginning of the war, Britain simply drew down massive stores of cotton. These stockpiles lasted until near the end of 1862. Why did Britain not recognize the Confederacy at that point? Two reasons: The Emancipation Proclamation and new sources of cotton. Having outlawed slavery throughout the United Kingdom in 1833, it was politically impossible for Great Britain, empty cotton warehouses or not, to recognize, diplomatically, the Confederate States. In addition, during the two years it took to draw down the stockpiles, Britain expanded cotton imports from India, Egypt, and Brazil.

Monopoly sellers often see no threats to their superior marketplace position. In these examples did the power of the monopoly blind the decision makers to other possibilities? Perhaps. As a result of their actions, this is how history unfolded.

KEY TERMS

allocative efficiency producing the optimal quantity of some output; the quantity where the marginal benefit to society of one more unit just equals the marginal cost

barriers to entry the legal, technological, or market forces that may discourage or prevent potential competitors from entering a market

copyright a form of legal protection to prevent copying, for commercial purposes, original works of authorship, including books and music

deregulation removing government controls over setting prices and quantities in certain industries

intellectual property the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their inventions

legal monopoly legal prohibitions against competition, such as regulated monopolies and intellectual property protection

marginal profit profit of one more unit of output, computed as marginal revenue minus marginal cost

monopoly a situation in which one firm produces all of the output in a market

natural monopoly economic conditions in the industry, for example, economies of scale or control of a critical resource, that limit effective competition

patent a government rule that gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time

predatory pricing when an existing firm uses sharp but temporary price cuts to discourage new competition

trade secrets methods of production kept secret by the producing firm

trademark an identifying symbol or name for a particular good and can only be used by the firm that registered that trademark

KEY CONCEPTS AND SUMMARY

9.1 How Monopolies Form: Barriers to Entry

Barriers to entry prevent or discourage competitors from entering the market. These barriers include: economies of scale that lead to natural monopoly; control of a physical resource; legal restrictions on competition; patent, trademark and copyright protection; and practices to intimidate the competition like predatory pricing. Intellectual property refers to legally guaranteed ownership of an idea, rather than a physical item. The laws that protect intellectual property include patents, copyrights, trademarks, and trade secrets. A natural monopoly arises when economies of scale persist over a large enough range of output that if one firm supplies the entire market, no other firm can enter without facing a cost disadvantage.

9.2 How a Profit-Maximizing Monopoly Chooses Output and Price

A monopolist is not a price taker, because when it decides what quantity to produce, it also determines the market price. For a monopolist, total revenue is relatively low at low quantities of output, because it is not selling much. Total revenue is also relatively low at very high quantities of output, because a very high quantity will sell only at a low price. Thus, total revenue for a monopolist will start low, rise, and then decline. The marginal revenue for a monopolist from selling additional units will decline. Each additional unit a monopolist sells will push down the overall market price, and as it sells more units, this lower price applies to increasingly more units.

The monopolist will select the profit-maximizing level of output where MR = MC, and then charge the price for that quantity of output as determined by the market demand curve. If that price is above average cost, the monopolist earns positive profits.

Monopolists are not productively efficient, because they do not produce at the minimum of the average cost curve. Monopolists are not allocatively efficient, because they do not produce at the quantity where P = MC. As a result, monopolists produce less, at a higher average cost, and charge a higher price than would a combination of firms in a perfectly competitive industry. Monopolists also may lack incentives for innovation, because they need not fear entry.

SELF-CHECK QUESTIONS

1. Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.

a. A patented invention
b. A popular but easily copied restaurant recipe
c. An industry where economies of scale are very small compared to the size of demand in the market
d. A well-established reputation for slashing prices in response to new entry
e. A well-respected brand name that has been carefully built up over many years

2. Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.

a. A city passes a law on how many licenses it will issue for taxicabs
b. A city passes a law that all taxicab drivers must pass a driving safety test and have insurance
c. A well-known trademark
d. Owning a spring that offers very pure water
e. An industry where economies of scale are very large compared to the size of demand in the market

3. Suppose the local electrical utility, a legal monopoly based on economies of scale, was split into four firms of equal size, with the idea that eliminating the monopoly would promote competitive pricing of electricity. What do you anticipate would happen to prices?

4. If Congress reduced the period of patent protection from 20 years to 10 years, what would likely happen to the amount of private research and development?

5. Suppose demand for a monopoly's product falls so that its profit-maximizing price is below average variable cost. How much output should the firm supply? *Hint*: Draw the graph.

6. Imagine a monopolist could charge a different price to every customer based on how much he or she were willing to pay. How would this affect monopoly profits?

REVIEW QUESTIONS

7. How is monopoly different from perfect competition?

8. What is a barrier to entry? Give some examples.

9. What is a natural monopoly?

10. What is a legal monopoly?

11. What is predatory pricing?

12. How is intellectual property different from other property?

13. What legal mechanisms protect intellectual property?

14. In what sense is a natural monopoly "natural"?

15. How is the demand curve perceived by a perfectly competitive firm different from the demand curve perceived by a monopolist?

16. How does the demand curve perceived by a monopolist compare with the market demand curve?

17. Is a monopolist a price taker? Explain briefly.

18. What is the usual shape of a total revenue curve for a monopolist? Why?

19. What is the usual shape of a marginal revenue curve for a monopolist? Why?

20. How can a monopolist identify the profit-maximizing level of output if it knows its total revenue and total cost curves?

21. How can a monopolist identify the profit-maximizing level of output if it knows its marginal revenue and marginal costs?

22. When a monopolist identifies its profit-maximizing quantity of output, how does it decide what price to charge?

23. Is a monopolist allocatively efficient? Why or why not?

24. How does the quantity produced and price charged by a monopolist compare to that of a perfectly competitive firm?

CRITICAL THINKING QUESTIONS

25. ALCOA does not have the monopoly power it once had. How do you suppose their barriers to entry were weakened?

26. Why are generic pharmaceuticals significantly cheaper than name brand ones?

27. For many years, the Justice Department has tried to break up large firms like IBM, Microsoft, and most recently Google, on the grounds that their large market share made them essentially monopolies. In a global market, where U.S. firms compete with firms from other countries, would this policy make the same sense as it might in a purely domestic context?

28. Intellectual property laws are intended to promote innovation, but some economists, such as Milton Friedman, have argued that such laws are not desirable. In the United States, there is no intellectual property protection for food recipes or for fashion designs. Considering the state of these two industries, and bearing in mind the discussion of the inefficiency of monopolies, can you think of any reasons why intellectual property laws might hinder innovation in some cases?

29. Imagine that you are managing a small firm and thinking about entering the market of a monopolist. The monopolist is currently charging a high price, and you have calculated that you can make a nice profit charging 10% less than the monopolist. Before you go ahead and challenge the monopolist, what possibility should you consider for how the monopolist might react?

30. If a monopoly firm is earning profits, how much would you expect these profits to be diminished by entry in the long run?

PROBLEMS

31. Return to **Figure 9.2**. Suppose P_0 is \$10 and P_1 is \$11. Suppose a new firm with the same LRAC curve as the incumbent tries to break into the market by selling 4,000 units of output. Estimate from the graph what the new firm's average cost of producing output would be. If the incumbent continues to produce 6,000 units, how much output would the two firms supply to the market? Estimate what would happen to the market price as a result of the supply of both the incumbent firm and the new entrant. Approximately how much profit would each firm earn?

32. Draw the demand curve, marginal revenue, and marginal cost curves from **Figure 9.6**, and identify the quantity of output the monopoly wishes to supply and the price it will charge. Suppose demand for the monopoly's product increases dramatically. Draw the new demand curve. What happens to the marginal revenue as a result of the increase in demand? What happens to the marginal cost curve? Identify the new profit-maximizing quantity and price. Does the answer make sense to you?

33. Draw a monopolist's demand curve, marginal revenue, and marginal cost curves. Identify the monopolist's profit-maximizing output level. Now, think about a slightly higher level of output (say $Q_0 + 1$). According to the graph, is there any consumer willing to pay more than the marginal cost of that new level of output? If so, what does this mean?

10 | Monopolistic Competition and Oligopoly

Figure 10.1 Competing Brands? The laundry detergent market is one that is characterized neither as perfect competition nor monopoly. (Credit: modification of work by Pixel Drip/Flickr Creative Commons)

Bring it Home

The Temptation to Defy the Law

Laundry detergent and bags of ice—products of industries that seem pretty mundane, maybe even boring. Hardly! Both have been the center of clandestine meetings and secret deals worthy of a spy novel. In France, between 1997 and 2004, the top four laundry detergent producers (Proctor & Gamble, Henkel, Unilever, and Colgate-Palmolive) controlled about 90 percent of the French soap market. Officials from the soap firms were meeting secretly, in out-of-the-way, small cafés around Paris. Their goals: Stamp out competition and set prices.

Around the same time, the top five Midwest ice makers (Home City Ice, Lang Ice, Tinley Ice, Sisler's Dairy, and Products of Ohio) had similar goals in mind when they secretly agreed to divide up the bagged ice market.

If both groups could meet their goals, it would enable each to act as though they were a single firm—in essence, a monopoly—and enjoy monopoly-size profits. The problem? In many parts of the world, including the European Union and the United States, it is illegal for firms to divide markets and set prices collaboratively.

These two cases provide examples of markets that are characterized neither as perfect competition nor monopoly. Instead, these firms are competing in market structures that lie between the extremes of monopoly

and perfect competition. How do they behave? Why do they exist? We will revisit this case later, to find out what happened.

Introduction to Monopolistic Competition and Oligopoly

In this chapter, you will learn about:

- Monopolistic Competition
- Oligopoly

Perfect competition and monopoly are at opposite ends of the competition spectrum. A perfectly competitive market has many firms selling identical products, who all act as price takers in the face of the competition. If you recall, price takers are firms that have no market power. They simply have to take the market price as given.

Monopoly arises when a single firm sells a product for which there are no close substitutes. We consider Microsoft, for instance, as a monopoly because it dominates the operating systems market.

What about the vast majority of real world firms and organizations that fall between these extremes, firms that we could describe as **imperfectly competitive**? What determines their behavior? They have more influence over the price they charge than perfectly competitive firms, but not as much as a monopoly. What will they do?

One type of imperfectly competitive market is **monopolistic competition**. Monopolistically competitive markets feature a large number of competing firms, but the products that they sell are not identical. Consider, as an example, the Mall of America in Minnesota, the largest shopping mall in the United States. In 2010, the Mall of America had 24 stores that sold women's "ready-to-wear" clothing (like Ann Taylor and Urban Outfitters), another 50 stores that sold clothing for both men and women (like Banana Republic, J. Crew, and Nordstrom's), plus 14 more stores that sold women's specialty clothing (like Motherhood Maternity and Victoria's Secret). Most of the markets that consumers encounter at the retail level are monopolistically competitive.

The other type of imperfectly competitive market is **oligopoly**. Oligopolistic markets are those which a small number of firms dominate. Commercial aircraft provides a good example: Boeing and Airbus each produce slightly less than 50% of the large commercial aircraft in the world. Another example is the U.S. soft drink industry, which Coca-Cola and Pepsi dominate. We characterize oligopolies by high barriers to entry with firms choosing output, pricing, and other decisions strategically based on the decisions of the other firms in the market. In this chapter, we first explore how monopolistically competitive firms will choose their profit-maximizing level of output. We will then discuss oligopolistic firms, which face two conflicting temptations: to collaborate as if they were a single monopoly, or to individually compete to gain profits by expanding output levels and cutting prices. Oligopolistic markets and firms can also take on elements of monopoly and of perfect competition.

10.1 | Monopolistic Competition

By the end of this section, you will be able to:

- Explain the significance of differentiated products
- Describe how a monopolistic competitor chooses price and quantity
- Discuss entry, exit, and efficiency as they pertain to monopolistic competition
- Analyze how advertising can impact monopolistic competition

Monopolistic competition involves many firms competing against each other, but selling products that are distinctive in some way. Examples include stores that sell different styles of clothing; restaurants or grocery stores that sell a variety of food; and even products like golf balls or beer that may be at least somewhat similar but differ in public perception because of advertising and brand names. There are over 600,000 restaurants in the United States. When products are distinctive, each firm has a mini-monopoly on its particular style or flavor or brand name. However, firms producing such products must also compete with other styles and flavors and brand names. The term "monopolistic competition" captures this mixture of mini-monopoly and tough competition, and the following Clear It Up feature

introduces its derivation.

Clear It Up

Who invented the theory of imperfect competition?

Two economists independently but simultaneously developed the theory of imperfect competition in 1933. The first was Edward Chamberlin of Harvard University who published *The Economics of Monopolistic Competition*. The second was Joan Robinson of Cambridge University who published *The Economics of Imperfect Competition*. Robinson subsequently became interested in macroeconomics and she became a prominent Keynesian, and later a post-Keynesian economist. (See the **Welcome to Economics!** and **The Keynesian Perspective (http://cnx.org/content/m63849/latest/)** chapters for more on Keynes.)

Differentiated Products

A firm can try to make its products different from those of its competitors in several ways: physical aspects of the product, location from which it sells the product, intangible aspects of the product, and perceptions of the product. We call products that are distinctive in one of these ways **differentiated products**.

Physical aspects of a product include all the phrases you hear in advertisements: unbreakable bottle, nonstick surface, freezer-to-microwave, non-shrink, extra spicy, newly redesigned for your comfort. A firm's location can also create a difference between producers. For example, a gas station located at a heavily traveled intersection can probably sell more gas, because more cars drive by that corner. A supplier to an automobile manufacturer may find that it is an advantage to locate close to the car factory.

Intangible aspects can differentiate a product, too. Some intangible aspects may be promises like a guarantee of satisfaction or money back, a reputation for high quality, services like free delivery, or offering a loan to purchase the product. Finally, product differentiation may occur in the minds of buyers. For example, many people could not tell the difference in taste between common varieties of ketchup or mayonnaise if they were blindfolded but, because of past habits and advertising, they have strong preferences for certain brands. Advertising can play a role in shaping these intangible preferences.

The concept of differentiated products is closely related to the degree of variety that is available. If everyone in the economy wore only blue jeans, ate only white bread, and drank only tap water, then the markets for clothing, food, and drink would be much closer to perfectly competitive. The variety of styles, flavors, locations, and characteristics creates product differentiation and monopolistic competition.

Perceived Demand for a Monopolistic Competitor

A monopolistically competitive firm perceives a demand for its goods that is an intermediate case between monopoly and competition. **Figure 10.2** offers a reminder that the demand curve that a perfectly competitive firm faces is perfectly elastic or flat, because the perfectly competitive firm can sell any quantity it wishes at the prevailing market price. In contrast, the demand curve, as faced by a monopolist, is the market demand curve, since a monopolist is the only firm in the market, and hence is downward sloping.

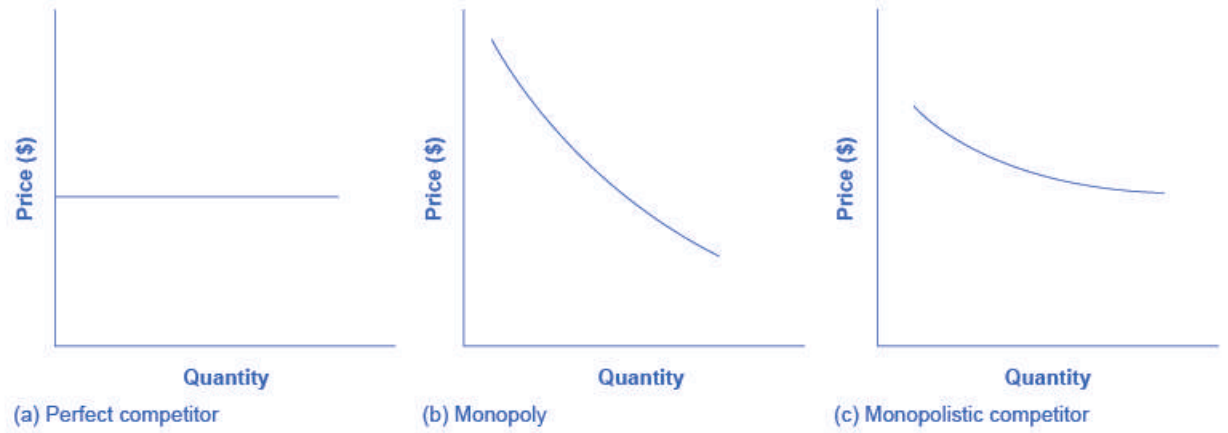


Figure 10.2 Perceived Demand for Firms in Different Competitive Settings The demand curve that a perfectly competitive firm faces is perfectly elastic, meaning it can sell all the output it wishes at the prevailing market price. The demand curve that a monopoly faces is the market demand. It can sell more output only by decreasing the price it charges. The demand curve that a monopolistically competitive firm faces falls in between.

The demand curve as a monopolistic competitor faces is not flat, but rather downward-sloping, which means that the monopolistic competitor can raise its price without losing all of its customers or lower the price and gain more customers. Since there are substitutes, the demand curve facing a monopolistically competitive firm is more elastic than that of a monopoly where there are no close substitutes. If a monopolist raises its price, some consumers will choose not to purchase its product—but they will then need to buy a completely different product. However, when a monopolistic competitor raises its price, some consumers will choose not to purchase the product at all, but others will choose to buy a similar product from another firm. If a monopolistic competitor raises its price, it will not lose as many customers as would a perfectly competitive firm, but it will lose more customers than would a monopoly that raised its prices.

At a glance, the demand curves that a monopoly and a monopolistic competitor face look similar—that is, they both slope down. However, the underlying economic meaning of these perceived demand curves is different, because a monopolist faces the market demand curve and a monopolistic competitor does not. Rather, a monopolistically competitive firm's demand curve is but one of many firms that make up the "before" market demand curve. Are you following? If so, how would you categorize the market for golf balls? Take a swing, then see the following Clear It Up feature.

Clear It Up

Are golf balls really differentiated products?

Monopolistic competition refers to an industry that has more than a few firms, each offering a product which, from the consumer's perspective, is different from its competitors. The U.S. Golf Association runs a laboratory that tests 20,000 golf balls a year. There are strict rules for what makes a golf ball legal. A ball's weight cannot exceed 1.620 ounces and its diameter cannot be less than 1.680 inches (which is a weight of 45.93 grams and a diameter of 42.67 millimeters, in case you were wondering). The Association also tests the balls by hitting them at different speeds. For example, the distance test involves having a mechanical golfer hit the ball with a titanium driver and a swing speed of 120 miles per hour. As the testing center explains: "The USGA system then uses an array of sensors that accurately measure the flight of a golf ball during a short, indoor trajectory from a ball launcher. From this flight data, a computer calculates the lift and drag forces that are generated by the speed, spin, and dimple pattern of the ball. ... The distance limit is 317 yards."

Over 1800 golf balls made by more than 100 companies meet the USGA standards. The balls do differ in various ways, such as the pattern of dimples on the ball, the types of plastic on the cover and in the cores, and other factors. Since all balls need to conform to the USGA tests, they are much more alike than different.

In other words, golf ball manufacturers are monopolistically competitive.

However, retail sales of golf balls are about $500 million per year, which means that many large companies have a powerful incentive to persuade players that golf balls are highly differentiated and that it makes a huge difference which one you choose. Sure, Tiger Woods can tell the difference. For the average amateur golfer who plays a few times a summer—and who loses many golf balls to the woods and lake and needs to buy new ones—most golf balls are pretty much indistinguishable.

How a Monopolistic Competitor Chooses Price and Quantity

The monopolistically competitive firm decides on its profit-maximizing quantity and price in much the same way as a monopolist. A monopolistic competitor, like a monopolist, faces a downward-sloping demand curve, and so it will choose some combination of price and quantity along its perceived demand curve.

As an example of a profit-maximizing monopolistic competitor, consider the Authentic Chinese Pizza store, which serves pizza with cheese, sweet and sour sauce, and your choice of vegetables and meats. Although Authentic Chinese Pizza must compete against other pizza businesses and restaurants, it has a differentiated product. The firm's perceived demand curve is downward sloping, as **Figure 10.3** shows and the first two columns of **Table 10.1**.

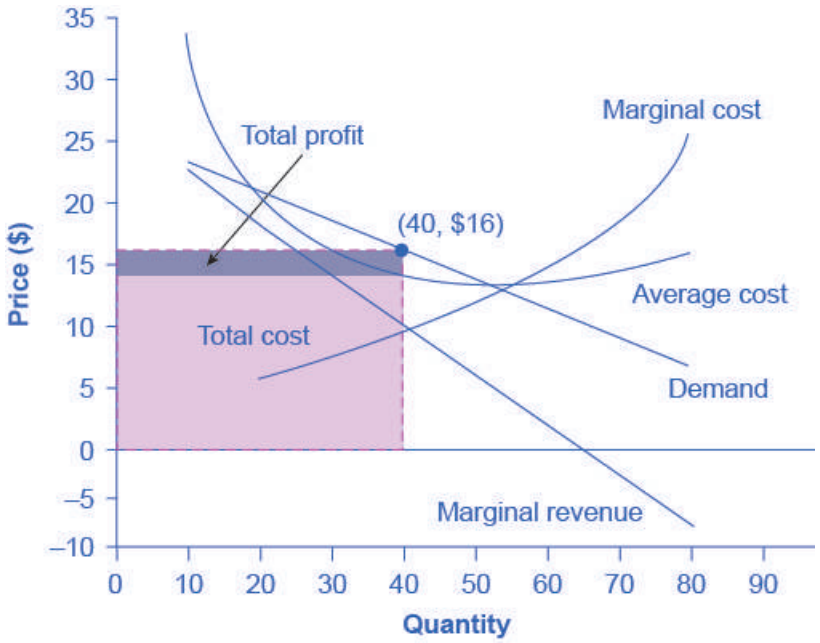


Figure 10.3 How a Monopolistic Competitor Chooses its Profit Maximizing Output and Price To maximize profits, the Authentic Chinese Pizza shop would choose a quantity where marginal revenue equals marginal cost, or Q where MR = MC. Here it would choose a quantity of 40 and a price of $16.

Quantity	Price	Total Revenue	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
10	$23	$230	$23	$340	$34	$34
20	$20	$400	$17	$400	$6	$20
30	$18	$540	$14	$480	$8	$16
40	$16	$640	$10	$580	$10	$14.50

Table 10.1 Revenue and Cost Schedule

Quantity	Price	Total Revenue	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
50	$14	$700	$6	$700	$12	$14
60	$12	$720	$2	$840	$14	$14
70	$10	$700	–$2	$1,020	$18	$14.57
80	$8	$640	–$6	$1,280	$26	$16

Table 10.1 Revenue and Cost Schedule

We can multiply the combinations of price and quantity at each point on the demand curve to calculate the total revenue that the firm would receive, which is in the third column of **Table 10.1**. We calculate marginal revenue, in the fourth column, as the change in total revenue divided by the change in quantity. The final columns of **Table 10.1** show total cost, marginal cost, and average cost. As always, we calculate marginal cost by dividing the change in total cost by the change in quantity, while we calculate average cost by dividing total cost by quantity. The following Work It Out feature shows how these firms calculate how much of their products to supply at what price.

Work It Out

How a Monopolistic Competitor Determines How Much to Produce and at What Price

The process by which a monopolistic competitor chooses its profit-maximizing quantity and price resembles closely how a monopoly makes these decisions process. First, the firm selects the profit-maximizing quantity to produce. Then the firm decides what price to charge for that quantity.

Step 1. The monopolistic competitor determines its profit-maximizing level of output. In this case, the Authentic Chinese Pizza company will determine the profit-maximizing quantity to produce by considering its marginal revenues and marginal costs. Two scenarios are possible:

- If the firm is producing at a quantity of output where marginal revenue exceeds marginal cost, then the firm should keep expanding production, because each marginal unit is adding to profit by bringing in more revenue than its cost. In this way, the firm will produce up to the quantity where MR = MC.
- If the firm is producing at a quantity where marginal costs exceed marginal revenue, then each marginal unit is costing more than the revenue it brings in, and the firm will increase its profits by reducing the quantity of output until MR = MC.

In this example, MR and MC intersect at a quantity of 40, which is the profit-maximizing level of output for the firm.

Step 2. The monopolistic competitor decides what price to charge. When the firm has determined its profit-maximizing quantity of output, it can then look to its perceived demand curve to find out what it can charge for that quantity of output. On the graph, we show this process as a vertical line reaching up through the profit-maximizing quantity until it hits the firm's perceived demand curve. For Authentic Chinese Pizza, it should charge a price of $16 per pizza for a quantity of 40.

Once the firm has chosen price and quantity, it's in a position to calculate total revenue, total cost, and profit. At a quantity of 40, the price of $16 lies above the average cost curve, so the firm is making economic profits. From **Table 10.1** we can see that, at an output of 40, the firm's total revenue is $640 and its total cost is $580, so profits are $60. In **Figure 10.3**, the firm's total revenues are the rectangle with the quantity of 40 on the horizontal axis and the price of $16 on the vertical axis. The firm's total costs are the light shaded rectangle with the same quantity of 40 on the horizontal axis but the average cost of $14.50 on the vertical axis. Profits are total revenues minus total costs, which is the shaded area above the average cost curve.

Although the process by which a monopolistic competitor makes decisions about quantity and price is similar to the way in which a monopolist makes such decisions, two differences are worth remembering. First, although both a monopolist and a monopolistic competitor face downward-sloping demand curves, the monopolist's perceived demand curve is the market demand curve, while the perceived demand curve for a monopolistic competitor is based on the extent of its product differentiation and how many competitors it faces. Second, a monopolist is surrounded by barriers to entry and need not fear entry, but a monopolistic competitor who earns profits must expect the entry of firms with similar, but differentiated, products.

Monopolistic Competitors and Entry

If one monopolistic competitor earns positive economic profits, other firms will be tempted to enter the market. A gas station with a great location must worry that other gas stations might open across the street or down the road—and perhaps the new gas stations will sell coffee or have a carwash or some other attraction to lure customers. A successful restaurant with a unique barbecue sauce must be concerned that other restaurants will try to copy the sauce or offer their own unique recipes. A laundry detergent with a great reputation for quality must take note that other competitors may seek to build their own reputations.

The entry of other firms into the same general market (like gas, restaurants, or detergent) shifts the demand curve that a monopolistically competitive firm faces. As more firms enter the market, the quantity demanded at a given price for any particular firm will decline, and the firm's perceived demand curve will shift to the left. As a firm's perceived demand curve shifts to the left, its marginal revenue curve will shift to the left, too. The shift in marginal revenue will change the profit-maximizing quantity that the firm chooses to produce, since marginal revenue will then equal marginal cost at a lower quantity.

Figure 10.4 (a) shows a situation in which a monopolistic competitor was earning a profit with its original perceived demand curve (D_0). The intersection of the marginal revenue curve (MR_0) and marginal cost curve (MC) occurs at point S, corresponding to quantity Q_0, which is associated on the demand curve at point T with price P_0. The combination of price P_0 and quantity Q_0 lies above the average cost curve, which shows that the firm is earning positive economic profits.

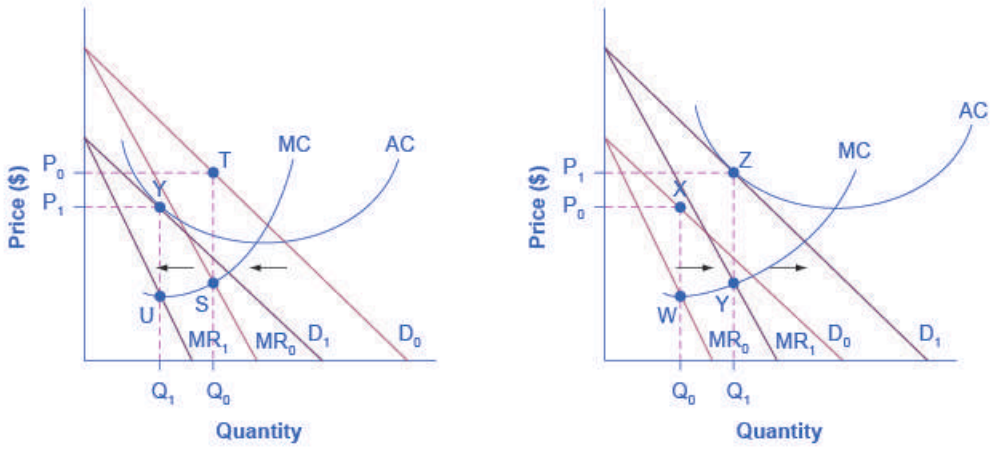


Figure 10.4 Monopolistic Competition, Entry, and Exit (a) At P_0 and Q_0, the monopolistically competitive firm in this figure is making a positive economic profit. This is clear because if you follow the dotted line above Q_0, you can see that price is above average cost. Positive economic profits attract competing firms to the industry, driving the original firm's demand down to D_1. At the new equilibrium quantity (P_1, Q_1), the original firm is earning zero economic profits, and entry into the industry ceases. In (b) the opposite occurs. At P_0 and Q_0, the firm is losing money. If you follow the dotted line above Q_0, you can see that average cost is above price. Losses induce firms to leave the industry. When they do, demand for the original firm rises to D_1, where once again the firm is earning zero economic profit.

Unlike a monopoly, with its high barriers to entry, a monopolistically competitive firm with positive economic profits will attract competition. When another competitor enters the market, the original firm's perceived demand curve shifts to the left, from D_0 to D_1, and the associated marginal revenue curve shifts from MR_0 to MR_1. The new profit-

maximizing output is Q_1, because the intersection of the MR_1 and MC now occurs at point U. Moving vertically up from that quantity on the new demand curve, the optimal price is at P_1.

As long as the firm is earning positive economic profits, new competitors will continue to enter the market, reducing the original firm's demand and marginal revenue curves. The long-run equilibrium is in the figure at point Y, where the firm's perceived demand curve touches the average cost curve. When price is equal to average cost, economic profits are zero. Thus, although a monopolistically competitive firm may earn positive economic profits in the short term, the process of new entry will drive down economic profits to zero in the long run. Remember that zero economic profit is not equivalent to zero accounting profit. A zero economic profit means the firm's accounting profit is equal to what its resources could earn in their next best use. **Figure 10.4** (b) shows the reverse situation, where a monopolistically competitive firm is originally losing money. The adjustment to long-run equilibrium is analogous to the previous example. The economic losses lead to firms exiting, which will result in increased demand for this particular firm, and consequently lower losses. Firms exit up to the point where there are no more losses in this market, for example when the demand curve touches the average cost curve, as in point Z.

Monopolistic competitors can make an economic profit or loss in the short run, but in the long run, entry and exit will drive these firms toward a zero economic profit outcome. However, the zero economic profit outcome in monopolistic competition looks different from the zero economic profit outcome in perfect competition in several ways relating both to efficiency and to variety in the market.

Monopolistic Competition and Efficiency

The long-term result of entry and exit in a perfectly competitive market is that all firms end up selling at the price level determined by the lowest point on the average cost curve. This outcome is why perfect competition displays productive efficiency: goods are produced at the lowest possible average cost. However, in monopolistic competition, the end result of entry and exit is that firms end up with a price that lies on the downward-sloping portion of the average cost curve, not at the very bottom of the AC curve. Thus, monopolistic competition will not be productively efficient.

In a perfectly competitive market, each firm produces at a quantity where price is set equal to marginal cost, both in the short and long run. This outcome is why perfect competition displays allocative efficiency: the social benefits of additional production, as measured by the marginal benefit, which is the same as the price, equal the marginal costs to society of that production. In a monopolistically competitive market, the rule for maximizing profit is to set MR = MC—and price is higher than marginal revenue, not equal to it because the demand curve is downward sloping. When P > MC, which is the outcome in a monopolistically competitive market, the benefits to society of providing additional quantity, as measured by the price that people are willing to pay, exceed the marginal costs to society of producing those units. A monopolistically competitive firm does not produce more, which means that society loses the net benefit of those extra units. This is the same argument we made about monopoly, but in this case the allocative inefficiency will be smaller. Thus, a monopolistically competitive industry will produce a lower quantity of a good and charge a higher price for it than would a perfectly competitive industry. See the following Clear It Up feature for more detail on the impact of demand shifts.

Why does a shift in perceived demand cause a shift in marginal revenue?

We use the combinations of price and quantity at each point on a firm's perceived demand curve to calculate total revenue for each combination of price and quantity. We then use this information on total revenue to calculate marginal revenue, which is the change in total revenue divided by the change in quantity. A change in perceived demand will change total revenue at every quantity of output and in turn, the change in total revenue will shift marginal revenue at each quantity of output. Thus, when entry occurs in a monopolistically competitive industry, the perceived demand curve for each firm will shift to the left, because a smaller quantity will be demanded at any given price. Another way of interpreting this shift in demand is to notice that, for each quantity sold, the firm will charge a lower price. Consequently, the marginal revenue will be lower for

each quantity sold—and the marginal revenue curve will shift to the left as well. Conversely, exit causes the perceived demand curve for a monopolistically competitive firm to shift to the right and the corresponding marginal revenue curve to shift right, too.

A monopolistically competitive industry does not display productive or allocative efficiency in either the short run, when firms are making economic profits and losses, nor in the long run, when firms are earning zero profits.

The Benefits of Variety and Product Differentiation

Even though monopolistic competition does not provide productive efficiency or allocative efficiency, it does have benefits of its own. Product differentiation is based on variety and innovation. Most people would prefer to live in an economy with many kinds of clothes, foods, and car styles; not in a world of perfect competition where everyone will always wear blue jeans and white shirts, eat only spaghetti with plain red sauce, and drive an identical model of car. Most people would prefer to live in an economy where firms are struggling to figure out ways of attracting customers by methods like friendlier service, free delivery, guarantees of quality, variations on existing products, and a better shopping experience.

Economists have struggled, with only partial success, to address the question of whether a market-oriented economy produces the optimal amount of variety. Critics of market-oriented economies argue that society does not really need dozens of different athletic shoes or breakfast cereals or automobiles. They argue that much of the cost of creating such a high degree of product differentiation, and then of advertising and marketing this differentiation, is socially wasteful—that is, most people would be just as happy with a smaller range of differentiated products produced and sold at a lower price. Defenders of a market-oriented economy respond that if people do not want to buy differentiated products or highly advertised brand names, no one is forcing them to do so. Moreover, they argue that consumers benefit substantially when firms seek short-term profits by providing differentiated products. This controversy may never be fully resolved, in part because deciding on the optimal amount of variety is very difficult, and in part because the two sides often place different values on what variety means for consumers. Read the following Clear It Up feature for a discussion on the role that advertising plays in monopolistic competition.

How does advertising impact monopolistic competition?

The U.S. economy spent about $180.12 billion on advertising in 2014, according to eMarketer.com. Roughly one third of this was television advertising, and another third was divided roughly equally between internet, newspapers, and radio. The remaining third was divided between direct mail, magazines, telephone directory yellow pages, and billboards. Mobile devices are increasing the opportunities for advertisers.

Advertising is all about explaining to people, or making people believe, that the products of one firm are differentiated from another firm's products. In the framework of monopolistic competition, there are two ways to conceive of how advertising works: either advertising causes a firm’s perceived demand curve to become more inelastic (that is, it causes the perceived demand curve to become steeper); or advertising causes demand for the firm’s product to increase (that is, it causes the firm’s perceived demand curve to shift to the right). In either case, a successful advertising campaign may allow a firm to sell either a greater quantity or to charge a higher price, or both, and thus increase its profits.

However, economists and business owners have also long suspected that much of the advertising may only offset other advertising. Economist A. C. Pigou wrote the following back in 1920 in his book, *The Economics of Welfare*:

> It may happen that expenditures on advertisement made by competing monopolists [that is, what we now call monopolistic competitors] will simply neutralise one another, and leave the industrial position exactly as it would have been if neither had expended anything. For, clearly, if each of two rivals makes equal efforts to attract the favour of the public away from the other, the total result is the same as it would have been if neither had made any effort at all.

10.2 | Oligopoly

By the end of this section, you will be able to:

- Explain why and how oligopolies exist
- Contrast collusion and competition
- Interpret and analyze the prisoner's dilemma diagram
- Evaluate the tradeoffs of imperfect competition

Many purchases that individuals make at the retail level are produced in markets that are neither perfectly competitive, monopolies, nor monopolistically competitive. Rather, they are oligopolies. Oligopoly arises when a small number of large firms have all or most of the sales in an industry. Examples of oligopoly abound and include the auto industry, cable television, and commercial air travel. Oligopolistic firms are like cats in a bag. They can either scratch each other to pieces or cuddle up and get comfortable with one another. If oligopolists compete hard, they may end up acting very much like perfect competitors, driving down costs and leading to zero profits for all. If oligopolists collude with each other, they may effectively act like a monopoly and succeed in pushing up prices and earning consistently high levels of profit. We typically characterize oligopolies by mutual interdependence where various decisions such as output, price, and advertising depend on other firm(s)' decisions. Analyzing the choices of oligopolistic firms about pricing and quantity produced involves considering the pros and cons of competition versus collusion at a given point in time.

Why Do Oligopolies Exist?

A combination of the barriers to entry that create monopolies and the product differentiation that characterizes monopolistic competition can create the setting for an oligopoly. For example, when a government grants a patent for an invention to one firm, it may create a monopoly. When the government grants patents to, for example, three different pharmaceutical companies that each has its own drug for reducing high blood pressure, those three firms may become an oligopoly.

Similarly, a natural monopoly will arise when the quantity demanded in a market is only large enough for a single firm to operate at the minimum of the long-run average cost curve. In such a setting, the market has room for only one firm, because no smaller firm can operate at a low enough average cost to compete, and no larger firm could sell what it produced given the quantity demanded in the market.

Quantity demanded in the market may also be two or three times the quantity needed to produce at the minimum of the average cost curve—which means that the market would have room for only two or three oligopoly firms (and they need not produce differentiated products). Again, smaller firms would have higher average costs and be unable to compete, while additional large firms would produce such a high quantity that they would not be able to sell it at a profitable price. This combination of economies of scale and market demand creates the barrier to entry, which led to the Boeing-Airbus oligopoly (also called a duopoly) for large passenger aircraft.

The product differentiation at the heart of monopolistic competition can also play a role in creating oligopoly. For example, firms may need to reach a certain minimum size before they are able to spend enough on advertising and marketing to create a recognizable brand name. The problem in competing with, say, Coca-Cola or Pepsi is not that producing fizzy drinks is technologically difficult, but rather that creating a brand name and marketing effort to equal Coke or Pepsi is an enormous task.

Collusion or Competition?

When oligopoly firms in a certain market decide what quantity to produce and what price to charge, they face a temptation to act as if they were a monopoly. By acting together, oligopolistic firms can hold down industry output, charge a higher price, and divide the profit among themselves. When firms act together in this way to reduce output and keep prices high, it is called **collusion**. A group of firms that have a formal agreement to collude to produce the monopoly output and sell at the monopoly price is called a **cartel**. See the following Clear It Up feature for a more in-depth analysis of the difference between the two.

Clear It Up

Collusion versus cartels: How to differentiate

In the United States, as well as many other countries, it is illegal for firms to collude since collusion is anti-competitive behavior, which is a violation of antitrust law. Both the Antitrust Division of the Justice Department and the Federal Trade Commission have responsibilities for preventing collusion in the United States.

The problem of enforcement is finding hard evidence of collusion. Cartels are formal agreements to collude. Because cartel agreements provide evidence of collusion, they are rare in the United States. Instead, most collusion is tacit, where firms implicitly reach an understanding that competition is bad for profits.

Economists have understood for a long time the desire of businesses to avoid competing so that they can instead raise the prices that they charge and earn higher profits. Adam Smith wrote in *Wealth of Nations* in 1776: "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices."

Even when oligopolists recognize that they would benefit as a group by acting like a monopoly, each individual oligopoly faces a private temptation to produce just a slightly higher quantity and earn slightly higher profit—while still counting on the other oligopolists to hold down their production and keep prices high. If at least some oligopolists give in to this temptation and start producing more, then the market price will fall. A small handful of oligopoly firms may end up competing so fiercely that they all find themselves earning zero economic profits—as if they were perfect competitors.

The Prisoner's Dilemma

Because of the complexity of oligopoly, which is the result of mutual interdependence among firms, there is no single, generally-accepted theory of how oligopolies behave, in the same way that we have theories for all the other market structures. Instead, economists use **game theory**, a branch of mathematics that analyzes situations in which players must make decisions and then receive payoffs based on what other players decide to do. Game theory has found widespread applications in the social sciences, as well as in business, law, and military strategy.

The **prisoner's dilemma** is a scenario in which the gains from cooperation are larger than the rewards from pursuing self-interest. It applies well to oligopoly. The story behind the prisoner's dilemma goes like this:

> Two co-conspiratorial criminals are arrested. When they are taken to the police station, they refuse to say anything and are put in separate interrogation rooms. Eventually, a police officer enters the room where Prisoner A is being held and says: "You know what? Your partner in the other room is confessing. Your partner is going to get a light prison sentence of just one year, and because you're remaining silent, the judge is going to stick you with eight years in prison. Why don't you get smart? If you confess, too, we'll cut your jail time down to five years, and your partner will get five years, also." Over in the next room, another police officer is giving exactly the same speech to Prisoner B. What the police officers do not say is that if both prisoners remain silent, the evidence against them is not especially strong, and the prisoners will end up with only two years in jail each.

The game theory situation facing the two prisoners is in **Table 10.2**. To understand the dilemma, first consider the choices from Prisoner A's point of view. If A believes that B will confess, then A should confess, too, so as to not get stuck with the eight years in prison. However, if A believes that B will not confess, then A will be tempted to act selfishly and confess, so as to serve only one year. The key point is that A has an incentive to confess regardless of what choice B makes! B faces the same set of choices, and thus will have an incentive to confess regardless of what choice A makes. To confess is called the dominant strategy. It is the strategy an individual (or firm) will pursue regardless of the other individual's (or firm's) decision. The result is that if prisoners pursue their own self-interest, both are likely to confess, and end up doing a total of 10 years of jail time between them.

		Prisoner B	
		Remain Silent (cooperate with other prisoner)	Confess (do not cooperate with other prisoner)
Prisoner A	Remain Silent (cooperate with other prisoner)	A gets 2 years, B gets 2 years	A gets 8 years, B gets 1 year
	Confess (do not cooperate with other prisoner)	A gets 1 year, B gets 8 years	A gets 5 years B gets 5 years

Table 10.2 The Prisoner's Dilemma Problem

The game is called a dilemma because if the two prisoners had cooperated by both remaining silent, they would only have had to serve a total of four years of jail time between them. If the two prisoners can work out some way of cooperating so that neither one will confess, they will both be better off than if they each follow their own individual self-interest, which in this case leads straight into longer jail terms.

The Oligopoly Version of the Prisoner's Dilemma

The members of an oligopoly can face a prisoner's dilemma, also. If each of the oligopolists cooperates in holding down output, then high monopoly profits are possible. Each oligopolist, however, must worry that while it is holding down output, other firms are taking advantage of the high price by raising output and earning higher profits. **Table 10.3** shows the prisoner's dilemma for a two-firm oligopoly—known as a **duopoly**. If Firms A and B both agree to hold down output, they are acting together as a monopoly and will each earn $1,000 in profits. However, both firms' dominant strategy is to increase output, in which case each will earn $400 in profits.

		Firm B	
		Hold Down Output (cooperate with other firm)	Increase Output (do not cooperate with other firm)
Firm A	Hold Down Output (cooperate with other firm)	A gets $1,000, B gets $1,000	A gets $200, B gets $1,500
	Increase Output (do not cooperate with other firm)	A gets $1,500, B gets $200	A gets $400, B gets $400

Table 10.3 A Prisoner's Dilemma for Oligopolists

Can the two firms trust each other? Consider the situation of Firm A:

- If A thinks that B will cheat on their agreement and increase output, then A will increase output, too, because for A the profit of $400 when both firms increase output (the bottom right-hand choice in **Table 10.3**) is better than a profit of only $200 if A keeps output low and B raises output (the upper right-hand choice in the table).
- If A thinks that B will cooperate by holding down output, then A may seize the opportunity to earn higher profits by raising output. After all, if B is going to hold down output, then A can earn $1,500 in profits by expanding output (the bottom left-hand choice in the table) compared with only $1,000 by holding down output as well (the upper left-hand choice in the table).

Thus, firm A will reason that it makes sense to expand output if B holds down output and that it also makes sense to expand output if B raises output. Again, B faces a parallel set of decisions that will lead B also to expand output.

The result of this prisoner's dilemma is often that even though A and B could make the highest combined profits by cooperating in producing a lower level of output and acting like a monopolist, the two firms may well end up in

a situation where they each increase output and earn only $400 each in profits. The following Clear It Up feature discusses one cartel scandal in particular.

Clear It Up

What is the Lysine cartel?

Lysine, a $600 million-a-year industry, is an amino acid that farmers use as a feed additive to ensure the proper growth of swine and poultry. The primary U.S. producer of lysine is Archer Daniels Midland (ADM), but several other large European and Japanese firms are also in this market. For a time in the first half of the 1990s, the world's major lysine producers met together in hotel conference rooms and decided exactly how much each firm would sell and what it would charge. The U.S. Federal Bureau of Investigation (FBI), however, had learned of the cartel and placed wire taps on a number of their phone calls and meetings.

From FBI surveillance tapes, following is a comment that Terry Wilson, president of the corn processing division at ADM, made to the other lysine producers at a 1994 meeting in Mona, Hawaii:

> I wanna go back and I wanna say something very simple. If we're going to trust each other, okay, and if I'm assured that I'm gonna get 67,000 tons by the year's end, we're gonna sell it at the prices we agreed to . . . The only thing we need to talk about there because we are gonna get manipulated by these [expletive] buyers—they can be smarter than us if we let them be smarter. . . . They [the customers] are not your friend. They are not my friend. And we gotta have 'em, but they are not my friends. You are my friend. I wanna be closer to you than I am to any customer. Cause you can make us ... money. ... And all I wanna tell you again is let's—let's put the prices on the board. Let's all agree that's what we're gonna do and then walk out of here and do it.

The price of lysine doubled while the cartel was in effect. Confronted by the FBI tapes, Archer Daniels Midland pled guilty in 1996 and paid a fine of $100 million. A number of top executives, both at ADM and other firms, later paid fines of up to $350,000 and were sentenced to 24–30 months in prison.

In another one of the FBI recordings, the president of Archer Daniels Midland told an executive from another competing firm that ADM had a slogan that, in his words, had "penetrated the whole company." The company president stated the slogan this way: "Our competitors are our friends. Our customers are the enemy." That slogan could stand as the motto of cartels everywhere.

How to Enforce Cooperation

How can parties who find themselves in a prisoner's dilemma situation avoid the undesired outcome and cooperate with each other? The way out of a prisoner's dilemma is to find a way to penalize those who do not cooperate.

Perhaps the easiest approach for colluding oligopolists, as you might imagine, would be to sign a contract with each other that they will hold output low and keep prices high. If a group of U.S. companies signed such a contract, however, it would be illegal. Certain international organizations, like the nations that are members of the Organization of Petroleum Exporting Countries (OPEC), have signed international agreements to act like a monopoly, hold down output, and keep prices high so that all of the countries can make high profits from oil exports. Such agreements, however, because they fall in a gray area of international law, are not legally enforceable. If Nigeria, for example, decides to start cutting prices and selling more oil, Saudi Arabia cannot sue Nigeria in court and force it to stop.

Link It Up

Visit the Organization of the Petroleum Exporting Countries website (http://openstaxcollege.org/l/OPEC) and learn more about its history and how it defines itself.

Because oligopolists cannot sign a legally enforceable contract to act like a monopoly, the firms may instead keep close tabs on what other firms are producing and charging. Alternatively, oligopolists may choose to act in a way that generates pressure on each firm to stick to its agreed quantity of output.

One example of the pressure these firms can exert on one another is the **kinked demand curve**, in which competing oligopoly firms commit to match price cuts, but not price increases. **Figure 10.5** shows this situation. Say that an oligopoly airline has agreed with the rest of a cartel to provide a quantity of 10,000 seats on the New York to Los Angeles route, at a price of $500. This choice defines the kink in the firm's perceived demand curve. The reason that the firm faces a kink in its demand curve is because of how the other oligopolists react to changes in the firm's price. If the oligopoly decides to produce more and cut its price, the other members of the cartel will immediately match any price cuts—and therefore, a lower price brings very little increase in quantity sold.

If one firm cuts its price to $300, it will be able to sell only 11,000 seats. However, if the airline seeks to raise prices, the other oligopolists will not raise their prices, and so the firm that raised prices will lose a considerable share of sales. For example, if the firm raises its price to $550, its sales drop to 5,000 seats sold. Thus, if oligopolists always match price cuts by other firms in the cartel, but do not match price increases, then none of the oligopolists will have a strong incentive to change prices, since the potential gains are minimal. This strategy can work like a silent form of cooperation, in which the cartel successfully manages to hold down output, increase price, and share a monopoly level of profits even without any legally enforceable agreement.

Figure 10.5 A Kinked Demand Curve Consider a member firm in an oligopoly cartel that is supposed to produce a quantity of 10,000 and sell at a price of $500. The other members of the cartel can encourage this firm to honor its commitments by acting so that the firm faces a kinked demand curve. If the oligopolist attempts to expand output and reduce price slightly, other firms also cut prices immediately—so if the firm expands output to 11,000, the price per unit falls dramatically, to $300. On the other side, if the oligopoly attempts to raise its price, other firms will not do so, so if the firm raises its price to $550, its sales decline sharply to 5,000. Thus, the members of a cartel can discipline each other to stick to the pre-agreed levels of quantity and price through a strategy of matching all price cuts but not matching any price increases.

Many real-world oligopolies, prodded by economic changes, legal and political pressures, and the egos of their top executives, go through episodes of cooperation and competition. If oligopolies could sustain cooperation with each

other on output and pricing, they could earn profits as if they were a single monopoly. However, each firm in an oligopoly has an incentive to produce more and grab a bigger share of the overall market; when firms start behaving in this way, the market outcome in terms of prices and quantity can be similar to that of a highly competitive market.

Tradeoffs of Imperfect Competition

Monopolistic competition is probably the single most common market structure in the U.S. economy. It provides powerful incentives for innovation, as firms seek to earn profits in the short run, while entry assures that firms do not earn economic profits in the long run. However, monopolistically competitive firms do not produce at the lowest point on their average cost curves. In addition, the endless search to impress consumers through product differentiation may lead to excessive social expenses on advertising and marketing.

Oligopoly is probably the second most common market structure. When oligopolies result from patented innovations or from taking advantage of economies of scale to produce at low average cost, they may provide considerable benefit to consumers. Oligopolies are often buffered by significant barriers to entry, which enable the oligopolists to earn sustained profits over long periods of time. Oligopolists also do not typically produce at the minimum of their average cost curves. When they lack vibrant competition, they may lack incentives to provide innovative products and high-quality service.

The task of public policy with regard to competition is to sort through these multiple realities, attempting to encourage behavior that is beneficial to the broader society and to discourage behavior that only adds to the profits of a few large companies, with no corresponding benefit to consumers. **Monopoly and Antitrust Policy** discusses the delicate judgments that go into this task.

Bring it Home

The Temptation to Defy the Law

Oligopolistic firms have been called “cats in a bag,” as this chapter mentioned. The French detergent makers chose to “cozy up” with each other. The result? An uneasy and tenuous relationship. When the *Wall Street Journal* reported on the matter, it wrote: “According to a statement a Henkel manager made to the [French anti-trust] commission, the detergent makers wanted ‘to limit the intensity of the competition between them and clean up the market.’ Nevertheless, by the early 1990s, a price war had broken out among them.” During the soap executives’ meetings, sometimes lasting more than four hours, the companies established complex pricing structures. “One [soap] executive recalled ‘chaotic’ meetings as each side tried to work out how the other had bent the rules.” Like many cartels, the soap cartel disintegrated due to the very strong temptation for each member to maximize its own individual profits.

How did this soap opera end? After an investigation, French antitrust authorities fined Colgate-Palmolive, Henkel, and Proctor & Gamble a total of €361 million ($484 million). A similar fate befell the icemakers. Bagged ice is a commodity, a perfect substitute, generally sold in 7- or 22-pound bags. No one cares what label is on the bag. By agreeing to carve up the ice market, control broad geographic swaths of territory, and set prices, the icemakers moved from perfect competition to a monopoly model. After the agreements, each firm was the sole supplier of bagged ice to a region. There were profits in both the long run and the short run. According to the courts: “These companies illegally conspired to manipulate the marketplace.” Fines totaled about $600,000—a steep fine considering a bag of ice sells for under $3 in most parts of the United States.

Even though it is illegal in many parts of the world for firms to set prices and carve up a market, the temptation to earn higher profits makes it extremely tempting to defy the law.

KEY TERMS

cartel a group of firms that collude to produce the monopoly output and sell at the monopoly price

collusion when firms act together to reduce output and keep prices high

differentiated product a product that is consumers perceive as distinctive in some way

duopoly an oligopoly with only two firms

game theory a branch of mathematics that economists use to analyze situations in which players must make decisions and then receive payoffs based on what decisions the other players make

imperfectly competitive firms and organizations that fall between the extremes of monopoly and perfect competition

kinked demand curve a perceived demand curve that arises when competing oligopoly firms commit to match price cuts, but not price increases

monopolistic competition many firms competing to sell similar but differentiated products

oligopoly when a few large firms have all or most of the sales in an industry

prisoner's dilemma a game in which the gains from cooperation are larger than the rewards from pursuing self-interest

product differentiation any action that firms do to make consumers think their products are different from their competitors'

KEY CONCEPTS AND SUMMARY

10.1 Monopolistic Competition

Monopolistic competition refers to a market where many firms sell differentiated products. Differentiated products can arise from characteristics of the good or service, location from which the firm sells the product, intangible aspects of the product, and perceptions of the product.

The perceived demand curve for a monopolistically competitive firm is downward-sloping, which shows that it is a price maker and chooses a combination of price and quantity. However, the perceived demand curve for a monopolistic competitor is more elastic than the perceived demand curve for a monopolist, because the monopolistic competitor has direct competition, unlike the pure monopolist. A profit-maximizing monopolistic competitor will seek out the quantity where marginal revenue is equal to marginal cost. The monopolistic competitor will produce that level of output and charge the price that the firm's demand curve indicates.

If the firms in a monopolistically competitive industry are earning economic profits, the industry will attract entry until profits are driven down to zero in the long run. If the firms in a monopolistically competitive industry are suffering economic losses, then the industry will experience exit of firms until economic losses are driven up to zero in the long run.

A monopolistically competitive firm is not productively efficient because it does not produce at the minimum of its average cost curve. A monopolistically competitive firm is not allocatively efficient because it does not produce where P = MC, but instead produces where P > MC. Thus, a monopolistically competitive firm will tend to produce a lower quantity at a higher cost and to charge a higher price than a perfectly competitive firm.

Monopolistically competitive industries do offer benefits to consumers in the form of greater variety and incentives for improved products and services. There is some controversy over whether a market-oriented economy generates too much variety.

10.2 Oligopoly

An oligopoly is a situation where a few firms sell most or all of the goods in a market. Oligopolists earn their

highest profits if they can band together as a cartel and act like a monopolist by reducing output and raising price. Since each member of the oligopoly can benefit individually from expanding output, such collusion often breaks down—especially since explicit collusion is illegal.

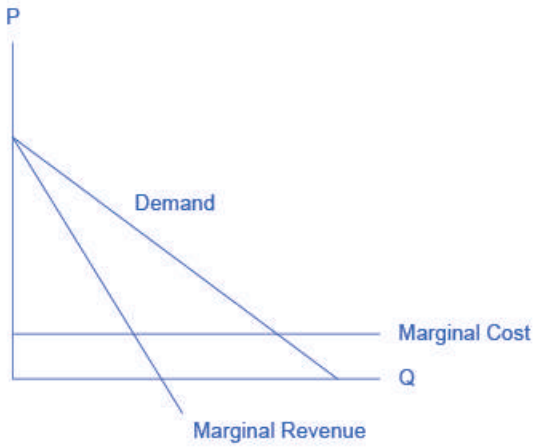
The prisoner's dilemma is an example of the application of game theory to analysis of oligopoly. It shows how, in certain situations, all sides can benefit from cooperative behavior rather than self-interested behavior. However, the challenge for the parties is to find ways to encourage cooperative behavior.

SELF-CHECK QUESTIONS

1. Suppose that, due to a successful advertising campaign, a monopolistic competitor experiences an increase in demand for its product. How will that affect the price it charges and the quantity it supplies?

2. Continuing with the scenario in question 1, in the long run, the positive economic profits that the monopolistic competitor earns will attract a response either from existing firms in the industry or firms outside. As those firms capture the original firm's profit, what will happen to the original firm's profit-maximizing price and output levels?

3. Consider the curve in **Figure 10.6**, which shows the market demand, marginal cost, and marginal revenue curve for firms in an oligopolistic industry. In this example, we assume firms have zero fixed costs.



a. Suppose the firms collude to form a cartel. What price will the cartel charge? What quantity will the cartel supply? How much profit will the cartel earn?
b. Suppose now that the cartel breaks up and the oligopolistic firms compete as vigorously as possible by cutting the price and increasing sales. What will be the industry quantity and price? What will be the collective profits of all firms in the industry?
c. Compare the equilibrium price, quantity, and profit for the cartel and cutthroat competition outcomes.

4. Sometimes oligopolies in the same industry are very different in size. Suppose we have a duopoly where one firm (Firm A) is large and the other firm (Firm B) is small, as the prisoner's dilemma box in **Table 10.4** shows.

	Firm B colludes with Firm A	Firm B cheats by selling more output
Firm A colludes with Firm B	A gets $1,000, B gets $100	A gets $800, B gets $200
Firm A cheats by selling more output	A gets $1,050, B gets $50	A gets $500, B gets $20

Table 10.4

Assuming that both firms know the payoffs, what is the likely outcome in this case?

REVIEW QUESTIONS

5. What is the relationship between product differentiation and monopolistic competition?

6. How is the perceived demand curve for a monopolistically competitive firm different from the perceived demand curve for a monopoly or a perfectly competitive firm?

7. How does a monopolistic competitor choose its profit-maximizing quantity of output and price?

8. How can a monopolistic competitor tell whether the price it is charging will cause the firm to earn profits or experience losses?

9. If the firms in a monopolistically competitive market are earning economic profits or losses in the short run, would you expect them to continue doing so in the long run? Why?

10. Is a monopolistically competitive firm productively efficient? Is it allocatively efficient? Why or why not?

11. Will the firms in an oligopoly act more like a monopoly or more like competitors? Briefly explain.

12. Does each individual in a prisoner's dilemma benefit more from cooperation or from pursuing self-interest? Explain briefly.

13. What stops oligopolists from acting together as a monopolist and earning the highest possible level of profits?

CRITICAL THINKING QUESTIONS

14. Aside from advertising, how can monopolistically competitive firms increase demand for their products?

15. Make a case for why monopolistically competitive industries never reach long-run equilibrium.

16. Would you rather have efficiency or variety? That is, one opportunity cost of the variety of products we have is that each product costs more per unit than if there were only one kind of product of a given type, like shoes. Perhaps a better question is, "What is the right amount of variety? Can there be too many varieties of shoes, for example?"

17. Would you expect the kinked demand curve to be more extreme (like a right angle) or less extreme (like a normal demand curve) if each firm in the cartel produces a near-identical product like OPEC and petroleum? What if each firm produces a somewhat different product? Explain your reasoning.

18. When OPEC raised the price of oil dramatically in the mid-1970s, experts said it was unlikely that the cartel could stay together over the long term—that the incentives for individual members to cheat would become too strong. More than forty years later, OPEC still exists. Why do you think OPEC has been able to beat the odds and continue to collude? *Hint:* You may wish to consider non-economic reasons.

PROBLEMS

19. Andrea's Day Spa began to offer a relaxing aromatherapy treatment. The firm asks you how much to charge to maximize profits. The first two columns in **Table 10.5** provide the price and quantity for the demand curve for treatments. The third column shows its total costs. For each level of output, calculate total revenue, marginal revenue, average cost, and marginal cost. What is the profit-maximizing level of output for the treatments and how much will the firm earn in profits?

Price	Quantity	TC
$25.00	0	$130
$24.00	10	$275
$23.00	20	$435
$22.50	30	$610
$22.00	40	$800
$21.60	50	$1,005
$21.20	60	$1,225

Table 10.5

20. Mary and Raj are the only two growers who provide organically grown corn to a local grocery store. They know that if they cooperated and produced less corn, they could raise the price of the corn. If they work independently, they will each earn $100. If they decide to work together and both lower their output, they can each earn $150. If one person lowers output and the other does not, the person who lowers output will earn $0 and the other person will capture the entire market and will earn $200. **Table 10.6** represents the choices available to Mary and Raj. What is the best choice for Raj if he is sure that Mary will cooperate? If Mary thinks Raj will cheat, what should Mary do and why? What is the prisoner's dilemma result? What is the preferred choice if they could ensure cooperation? A = Work independently; B = Cooperate and Lower Output. (Each results entry lists Raj's earnings first, and Mary's earnings second.)

		Mary	
		A	B
Raj	A	($100, $100)	($200, $0)
	B	($0, $200)	($150, $150)

Table 10.6

21. Jane and Bill are apprehended for a bank robbery. They are taken into separate rooms and questioned by the police about their involvement in the crime. The police tell them each that if they confess and turn the other person in, they will receive a lighter sentence. If they both confess, they will be each be sentenced to 30 years. If neither confesses, they will each receive a 20-year sentence. If only one confesses, the confessor will receive 15 years and the one who stayed silent will receive 35 years. **Table 10.7** below represents the choices available to Jane and Bill. If Jane trusts Bill to stay silent, what should she do? If Jane thinks that Bill will confess, what should she do? Does Jane have a dominant strategy? Does Bill have a dominant strategy? A = Confess; B = Stay Silent. (Each results entry lists Jane's sentence first (in years), and Bill's sentence second.)

		Jane	
		A	B
Bill	A	(30, 30)	(15, 35)
	B	(35, 15)	(20, 20)

Table 10.7

11 | Monopoly and Antitrust Policy

Figure 11.1 Oligopoly versus Competitors in the Marketplace Large corporations, such as the natural gas producer Kinder Morgan, can bring economies of scale to the marketplace. Will that benefit consumers, or is more competition better? (Credit: modification of work by Derrick Coetzee/Flickr Creative Commons)

Bring it Home

More than Cooking, Heating, and Cooling

If you live in the United States, there is a slightly better than 50–50 chance your home is heated and cooled using natural gas. You may even use natural gas for cooking. However, those uses are not the primary uses of natural gas in the U.S. In 2016, according to the U.S. Energy Information Administration, home heating, cooling, and cooking accounted for just 16% of natural gas usage. What accounts for the rest? The greatest uses for natural gas are the generation of electric power (36%) and in industry (28%). Together these three uses for natural gas touch many areas of our lives, so why would there be any opposition to a merger of two natural gas firms? After all, a merger could mean increased efficiencies and reduced costs to people like you and me.

In October 2011, Kinder Morgan and El Paso Corporation, two natural gas firms, announced they were merging. The announcement stated the combined firm would link “nearly every major production region with markets,” cut costs by “eliminating duplication in pipelines and other assets,” and that “the savings could be passed on to consumers.”

The objection? The $21.1 billion deal would give Kinder Morgan control of more than 80,000 miles of pipeline, making the new firm the third largest energy producer in North America. Policymakers and the public wondered whether the new conglomerate really would pass on cost savings to consumers, or would the merger give Kinder Morgan a strong oligopoly position in the natural gas marketplace?

That brings us to the central questions this chapter poses: What should the balance be between corporate size and a larger number of competitors in a marketplace, and what role should the government play in this balancing act?

Introduction to Monopoly and Antitrust Policy

In this chapter, you will learn about:

- Corporate Mergers
- Regulating Anticompetitive Behavior
- Regulating Natural Monopolies
- The Great Deregulation Experiment

The previous chapters on the theory of the firm identified three important lessons: First, that competition, by providing consumers with lower prices and a variety of innovative products, is a good thing; second, that large-scale production can dramatically lower average costs; and third, that markets in the real world are rarely perfectly competitive. As a consequence, government policymakers must determine how much to intervene to balance the potential benefits of large-scale production against the potential loss of competition that can occur when businesses grow in size, especially through mergers.

For example, in 2006, AT&T and BellSouth proposed a merger. At the time, there were very few mobile phone service providers. Both the Justice Department and the FCC blocked the proposal.

The two companies argued that the merger would benefit consumers, who would be able to purchase better telecommunications services at a cheaper price because the newly created firm would take advantage of economies of scale and eliminate duplicate investments. However, a number of activist groups like the Consumer Federation of America and Public Knowledge expressed fears that the merger would reduce competition and lead to higher prices for consumers for decades to come. In December 2006, the federal government allowed the merger to proceed. By 2009, the new post-merger AT&T was the eighth largest company by revenues in the United States, and by that measure the largest telecommunications company in the world. Economists have spent – and will still spend – years trying to determine whether the merger of AT&T and BellSouth, as well as other smaller mergers of telecommunications companies at about this same time, helped consumers, hurt them, or did not make much difference.

This chapter discusses public policy issues about competition. How can economists and governments determine when mergers of large companies like AT&T and BellSouth should be allowed and when they should be blocked? The government also plays a role in policing anticompetitive behavior other than mergers, like prohibiting certain kinds of contracts that might restrict competition. In the case of natural monopoly, however, trying to preserve competition probably will not work very well, and so government will often resort to regulation of price and/or quantity of output. In recent decades, there has been a global trend toward less government intervention in the price and output decisions of businesses.

11.1 | Corporate Mergers

By the end of this section, you will be able to:

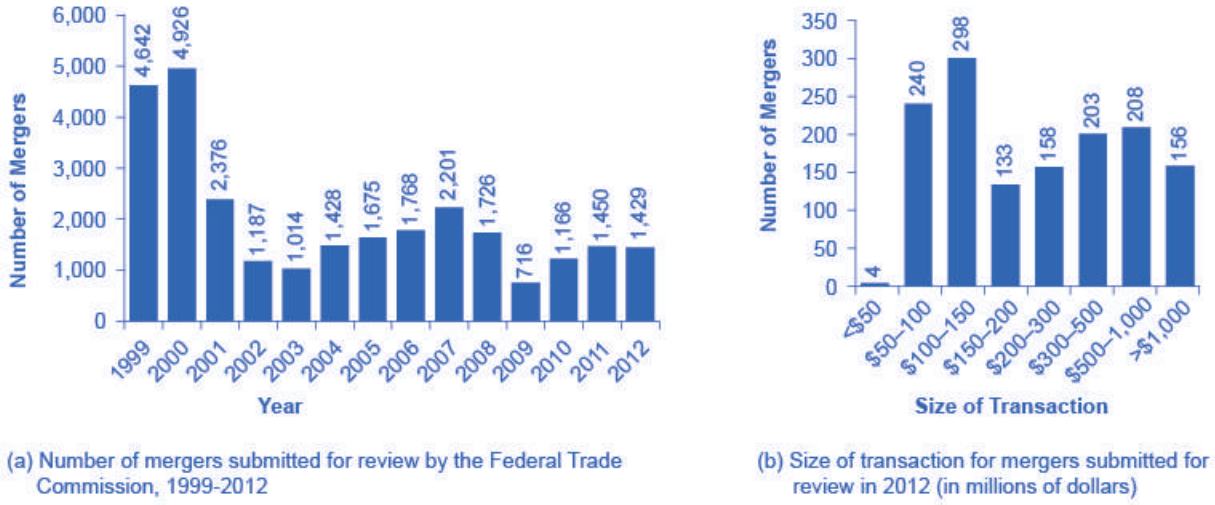
- Explain antitrust law and its significance
- Calculate concentration ratios
- Calculate the Herfindahl-Herschman Index (HHI)
- Evaluate methods of antitrust regulation

A corporate **merger** occurs when two formerly separate firms combine to become a single firm. When one firm purchases another, it is called an **acquisition**. An acquisition may not look just like a merger, since the newly purchased firm may continue to operate under its former company name. Mergers can also be lateral, where two firms of similar sizes combine to become one. However, both mergers and acquisitions lead to two formerly separate firms operating under common ownership, and so they are commonly grouped together.

Regulations for Approving Mergers

Since a merger combines two firms into one, it can reduce the extent of competition between firms. Therefore, when

two U.S. firms announce a merger or acquisition where at least one of the firms is above a minimum size of sales (a threshold that moves up gradually over time, and was at $76.3 million in 2015), or certain other conditions are met, they are required under law to notify the U.S. Federal Trade Commission (FTC). The left-hand panel of **Figure 11.2** (a) shows the number of mergers submitted for review to the FTC each year from 2002 to 2015. Mergers follow the business cycle, falling after the 2001 recession, peaking in 2007 as the Great Recession struck, and then rising since 2009. The right-hand panel of **Figure 11.2** (b) shows the distribution of those mergers submitted for review in 2015 as measured by the size of the transaction. It is important to remember that this total leaves out many small mergers under $50 million, which companies only need to report in certain limited circumstances. About a third of all reported merger and acquisition transactions in 2015 exceeded $500 million, while about 15 percent exceeded $1 billion.



(a) Number of mergers submitted for review by the Federal Trade Commission, 1999-2012

(b) Size of transaction for mergers submitted for review in 2012 (in millions of dollars)

Figure 11.2 Number and Size of Mergers (a) The number of mergers grew from 2003 to 2007, then fell dramatically during the 2008-2009 Great Recession, before recovering since. (b) In 2015, the greatest number of mergers submitted for review by the Federal Trade Commission was for transactions between $500 million and $1 billion.

The laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones, are called **antitrust laws**. Before a large merger happens, the antitrust regulators at the FTC and the U.S. Department of Justice can allow the merger, prohibit it, or allow it if certain conditions are met. One common condition is that the merger will be allowed if the firm agrees to sell off certain parts. For example, in 2006, Johnson & Johnson bought the Pfizer's "consumer health" division, which included well-known brands like Listerine mouthwash and Sudafed cold medicine. As a condition of allowing the merger, Johnson & Johnson was required to sell off six brands to other firms, including Zantac® heartburn relief medication, Cortizone anti-itch cream, and Balmex diaper rash medication, to preserve a greater degree of competition in these markets.

The U.S. government approves most proposed mergers. In a market-oriented economy, firms have the freedom to make their own choices. Private firms generally have the freedom to:

- expand or reduce production
- set the price they choose
- open new factories or sales facilities or close them
- hire workers or to lay them off
- start selling new products or stop selling existing ones

If the owners want to acquire a firm or be acquired, or to merge with another firm, this decision is just one of many that firms are free to make. In these conditions, the managers of private firms will sometimes make mistakes. They may close down a factory which, it later turns out, would have been profitable. They may start selling a product that ends up losing money. A merger between two companies can sometimes lead to a clash of corporate personalities that makes both firms worse off. However, the fundamental belief behind a market-oriented economy is that firms, not governments, are in the best position to know if their actions will lead to attracting more customers or producing

more efficiently.

Government regulators agree that most mergers are beneficial to consumers. As the Federal Trade Commission has noted on its website (as of November, 2013): “Most mergers actually benefit competition and consumers by allowing firms to operate more efficiently.” At the same time, the FTC recognizes, “Some [mergers] are likely to lessen competition. That, in turn, can lead to higher prices, reduced availability of goods or services, lower quality of products, and less innovation. Some mergers create a concentrated market, while others enable a single firm to raise prices.” The challenge for the antitrust regulators at the FTC and the U.S. Department of Justice is to figure out when a merger may hinder competition. This decision involves both numerical tools and some judgments that are difficult to quantify. The following Clear It Up explains the origins of U.S. antitrust law.

Clear It Up

What is U.S. antitrust law?

In the closing decades of the 1800s, many industries in the U.S. economy were dominated by a single firm that had most of the sales for the entire country. Supporters of these large firms argued that they could take advantage of economies of scale and careful planning to provide consumers with products at low prices. However, critics pointed out that when competition was reduced, these firms were free to charge more and make permanently higher profits, and that without the goading of competition, it was not clear that they were as efficient or innovative as they could be.

In many cases, these large firms were organized in the legal form of a “trust,” in which a group of formerly independent firms were consolidated by mergers and purchases, and a group of “trustees” then ran the companies as if they were a single firm. Thus, when the U.S. government sought to limit the power of these trusts, it passed the **Sherman Antitrust Act** in 1890 - the nation's first antitrust law. In an early demonstration of the law’s power, the U.S. Supreme Court in 1911 upheld the government’s right to break up Standard Oil, which had controlled about 90% of the country’s oil refining, into 34 independent firms, including Exxon, Mobil, Amoco, and Chevron. In 1914, the **Clayton Antitrust Act** outlawed mergers and acquisitions (where the outcome would be to “substantially lessen competition” in an industry), price discrimination (where different customers are charged different prices for the same product), and tied sales (where purchase of one product commits the buyer to purchase some other product). Also in 1914, the Federal Trade Commission (FTC) was created to define more specifically what competition was unfair. In 1950, the **Celler-Kefauver Act** extended the Clayton Act by restricting vertical and conglomerate mergers. A vertical merger occurs when two or more firms, operating at different levels within an industry's supply chain, merge operations. A conglomerate merger is a merger between firms that are involved in totally unrelated business activities. In the twenty-first century, the FTC and the U.S. Department of Justice continue to enforce antitrust laws.

The Four-Firm Concentration Ratio

Regulators have struggled for decades to measure the degree of monopoly power in an industry. An early tool was the **concentration ratio**, which measures the combined market share (or percent of total industry sales) which account for the largest firms (typically the top four to eight). For an explanation of how high market concentrations can create inefficiencies in an economy, refer to **Monopoly**.

Say that the market for replacing broken automobile windshields in a certain city has 18 firms with the market shares in **Table 11.1**, where the **market share** is each firm’s proportion of total sales in that market. We calculate the four-firm concentration ratio by adding the market shares of the four largest firms: in this case, 16 + 10 + 8 + 6 = 40. We do not consider this concentration ratio especially high, because the largest four firms have less than half the market.

If the market shares for replacing automobile windshields are:	
Smooth as Glass Repair Company	16% of the market

Table 11.1 Calculating Concentration Ratios from Market Shares

If the market shares for replacing automobile windshields are:	
The Auto Glass Doctor Company	10% of the market
Your Car Shield Company	8% of the market
Seven firms that each have 6% of the market	42% of the market, combined
Eight firms that each have 3% of the market	24% of the market, combined
Then the four-firm concentration ratio is 16 + 10 + 8 + 6 = 40.	

Table 11.1 Calculating Concentration Ratios from Market Shares

The concentration ratio approach can help to clarify some of the fuzziness over deciding when a merger might affect competition. For instance, if two of the smallest firms in the hypothetical market for repairing automobile windshields merged, the four-firm concentration ratio would not change—which implies that there is not much worry that the degree of competition in the market has notably diminished. However, if the top two firms merged, then the four-firm concentration ratio would become 46 (that is, 26 + 8 + 6 + 6). While this concentration ratio is modestly higher, the four-firm concentration ratio would still be less than half, so such a proposed merger might barely raise an eyebrow among antitrust regulators.

Link It Up

Visit this website (http://openstaxcollege.org/l/Google_FTC) to read an article about Google's run-in with the FTC.

The Herfindahl-Hirshman Index

A four-firm concentration ratio is a simple tool, which may reveal only part of the story. For example, consider two industries that both have a four-firm concentration ratio of 80. However, in one industry five firms each control 20% of the market, while in the other industry, the top firm holds 77% of the market and all the other firms have 1% each. Although the four-firm concentration ratios are identical, it would be reasonable to worry more about the extent of competition in the second case—where the largest firm is nearly a monopoly—than in the first.

Another approach to measuring industry concentration that can distinguish between these two cases is called the **Herfindahl-Hirschman Index (HHI)**. We calculate HHI by summing the squares of the market share of each firm in the industry, as the following Work It Out shows.

Work It Out

Calculating HHI

Step 1. Calculate the HHI for a monopoly with a market share of 100%. Because there is only one firm, it has 100% market share. The HHI is $100^2 = 10{,}000$.

Step 2. For an extremely competitive industry, with dozens or hundreds of extremely small competitors, the HHI value might drop as low as 100 or even less. Calculate the HHI for an industry with 100 firms that each have 1% of the market. In this case, the HHI is $100(1^2) = 100$.

Step 3. Calculate the HHI for the industry in **Table 11.1**. In this case, the HHI is $16^2 + 10^2 + 8^2 + 7(6^2) + 8(3^2) = 744$.

Step 4. Note that the HHI gives greater weight to large firms.

Step 5. Consider the earlier example, comparing one industry where five firms each have 20% of the market with an industry where one firm has 77% and the other 23 firms have 1% each. The two industries have the same four-firm concentration ratio of 80. However, the HHI for the first industry is $5(20^2) = 2{,}000$, while the HHI for the second industry is much higher at $77^2 + 23(1^2) = 5{,}952$.

Step 6. Note that the near-monopolist in the second industry drives up the HHI measure of industrial concentration.

Step 7. Review **Table 11.2** which gives some examples of the four-firm concentration ratio and the HHI in various U.S. industries in 2016. (You can find market share data from multiple industry sources. Data in the table are from: Statista.com (for wireless), *The Wall Street Journal* (for automobiles), Gartner.com (for computers) and the U.S. Bureau of Transportation Statistics (for airlines).)

U.S. Industry	Four-Firm Ratio	HHI
Wireless	98	2,736
Largest five: Verizon, AT&T, Sprint, T-Mobile, US Cellular		
Personal Computers	76	1,234
Largest five: HP, Lenovo, Dell, Asus, Apple, Acer		
Airlines	69	1,382
Largest five: American, Southwest, Delta, United, JetBlue		
Automobiles	58	1,099
Largest five: Ford, GM, Toyota, Chrysler, Nissan		

Table 11.2 Examples of Concentration Ratios and HHIs in the U.S. Economy, 2016

In the 1980s, the FTC followed these guidelines: If a merger would result in an HHI of less than 1,000, the FTC would probably approve it. If a merger would result in an HHI of more than 1,800, the FTC would probably challenge it. If a merger would result in an HHI between 1,000 and 1,800, then the FTC would scrutinize the plan and make a case-by-case decision. However, in the last several decades, the antitrust enforcement authorities have moved away from relying as heavily on measures of concentration ratios and HHIs to determine whether they will allow a merger, and instead they carry out more case-by-case analysis on the extent of competition in different industries.

New Directions for Antitrust

Both the four-firm concentration ratio and the Herfindahl-Hirschman index share some weaknesses. First, they begin

from the assumption that the "market" under discussion is well-defined, and the only question is measuring how sales are divided in that market. Second, they are based on an implicit assumption that competitive conditions across industries are similar enough that a broad measure of concentration in the market is enough to make a decision about the effects of a merger. These assumptions, however, are not always correct. In response to these two problems, the antitrust regulators have been changing their approach in the last decade or two.

Defining a **market** is often controversial. For example, Microsoft in the early 2000s had a dominant share of the software for computer operating systems. However, in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. A narrowly defined market will tend to make concentration appear higher, while a broadly defined market will tend to make it appear smaller.

In recent decades, there have been two especially important shifts affecting how we define markets: one centers on technology and the other centers on globalization. In addition, these two shifts are interconnected. With the vast improvement in communications technologies, including the development of the internet, a consumer can order books or pet supplies from all over the country or the world. As a result, the degree of competition many local retail businesses face has increased. The same effect may operate even more strongly in markets for business supplies, where so-called "business-to-business" websites can allow buyers and suppliers from anywhere in the world to find each other.

Globalization has changed the market boundaries. As recently as the 1970s, it was common for measurements of concentration ratios and HHIs to stop at national borders. Now, many industries find that their competition comes from the global market. A few decades ago, three companies, General Motors, Ford, and Chrysler, dominated the U.S. auto market. By 2014, however, production of these three firms accounted for less than half of U.S. auto sales, and they were facing competition from well-known car manufacturers such as Toyota, Honda, Nissan, Volkswagen, Mitsubishi, and Mazda. When analysts calculate HHIs with a global perspective, concentration in most major industries—including cars—is lower than in a purely domestic context.

Because attempting to define a particular market can be difficult and controversial, the Federal Trade Commission has begun to look less at market share and more at the data on actual competition between businesses. For example, in February 2007, Whole Foods Market and Wild Oats Market announced that they wished to merge. These were the two largest companies in the market that the government defined as "premium natural and organic supermarket chains." However, one could also argue that they were two relatively small companies in the broader market for all stores that sell groceries or specialty food products.

Rather than relying on a market definition, the government antitrust regulators looked at detailed evidence on profits and prices for specific stores in different cities, both before and after other competitive stores entered or exited. Based on that evidence, the Federal Trade Commission decided to block the merger. After two years of legal battles, the FTC eventually allowed the merger in 2009 under the conditions that Whole Foods sell off the Wild Oats brand name and a number of individual stores, to preserve competition in certain local markets. For more on the difficulties of defining markets, refer to **Monopoly**.

This new approach to antitrust regulation involves detailed analysis of specific markets and companies, instead of defining a market and counting up total sales. A common starting point is for antitrust regulators to use statistical tools and real-world evidence to estimate the **demand curves** and **supply curves** the firms proposing a merger face. A second step is to specify how competition occurs in this specific industry. Some possibilities include competing to cut prices, to raise output, to build a brand name through advertising, and to build a reputation for good service or high quality. With these pieces of the puzzle in place, it is then possible to build a statistical model that estimates the likely outcome for consumers if the two firms are allowed to merge. These models do require some degree of subjective judgment, and so they can become the subject of legal disputes between the antitrust authorities and the companies that wish to merge.

11.2 | Regulating Anticompetitive Behavior

By the end of this section, you will be able to:

- Analyze restrictive practices
- Explain tying sales, bundling, and predatory pricing
- Evaluate a real-world situation of possible anticompetitive and restrictive practices

The U.S. antitrust laws reach beyond blocking mergers that would reduce competition to include a wide array of anticompetitive practices. For example, it is illegal for competitors to form a cartel to collude to make pricing and output decisions, as if they were a monopoly firm. The Federal Trade Commission and the U.S. Department of Justice prohibit firms from agreeing to fix prices or output, rigging bids, or sharing or dividing markets by allocating customers, suppliers, territories, or lines of commerce.

In the late 1990s, for example, the antitrust regulators prosecuted an international cartel of vitamin manufacturers, including the Swiss firm Hoffman-La Roche, the German firm BASF, and the French firm Rhone-Poulenc. These firms reached agreements on how much to produce, how much to charge, and which firm would sell to which customers. Firms bought the high-priced vitamins like General Mills, Kellogg, Purina-Mills, and Proctor and Gamble which pushed up the prices more. Hoffman-La Roche pleaded guilty in May 1999 and agreed both to pay a fine of $500 million and to have at least one top executive serve four months of jail time.

Under U.S. antitrust laws, monopoly itself is not illegal. If a firm has a monopoly because of a newly patented invention, for example, the law explicitly allows a firm to earn higher-than-normal profits for a time as a reward for innovation. If a firm achieves a large share of the market by producing a better product at a lower price, such behavior is not prohibited by antitrust law.

Restrictive Practices

Antitrust law includes rules against **restrictive practices**—practices that do not involve outright agreements to raise price or to reduce the quantity produced, but that might have the effect of reducing competition. Antitrust cases involving restrictive practices are often controversial, because they delve into specific contracts or agreements between firms that are allowed in some cases but not in others.

For example, if a product manufacturer is selling to a group of dealers who then sell to the general public it is illegal for the manufacturer to demand a **minimum resale price maintenance agreement**, which would require the dealers to sell for at least a certain minimum price. A minimum price contract is illegal because it would restrict competition among dealers. However, the manufacturer is legally allowed to “suggest” minimum prices and to stop selling to dealers who regularly undercut the suggested price. If you think this rule sounds like a fairly subtle distinction, you are right.

An **exclusive dealing** agreement between a manufacturer and a dealer can be legal or illegal. It is legal if the purpose of the contract is to encourage competition between dealers. For example, it is legal for the Ford Motor Company to sell its cars to only Ford dealers, and for General Motors to sell to only GM dealers, and so on. However, exclusive deals may also limit competition. If one large retailer obtained the exclusive rights to be the sole distributor of televisions, computers, and audio equipment made by a number of companies, then this exclusive contract would have an anticompetitive effect on other retailers.

Tying sales happen when a customer is required to buy one product only if the customer also buys a second product. Tying sales are controversial because they force consumers to purchase a product that they may not actually want or need. Further, the additional, required products are not necessarily advantageous to the customer. Suppose that to purchase a popular DVD, the store required that you also purchase a certain portable TV model. These products are only loosely related, thus there is no reason to make the purchase of one contingent on the other. Even if a customer were interested in a portable TV, the tying to a particular model prevents the customer from having the option of selecting one from the numerous types available in the market.

A related, but not identical, concept is **bundling**, where a firm sells two or more products as one. Bundling typically offers an advantage for consumers by allowing them to acquire multiple products or services for a better price. For example, several cable companies allow customers to buy products like cable, internet, and a phone line through a special price available through bundling. Customers are also welcome to purchase these products separately, but the

price of bundling is usually more appealing.

In some cases, we can view tying sales and bundling as anticompetitive. However, in other cases they may be legal and even common. It is common for people to purchase season tickets to a sports team or a set of concerts so to guarantee tickets to the few contests or shows that are most popular and likely to sell out. Computer software manufacturers may often bundle a number of different programs, even when the buyer wants only a few. Think about the software that is included in a new computer purchase, for example.

Recall from the chapter on **Monopoly** that predatory pricing occurs when the existing firm (or firms) reacts to a new firm by dropping prices very low, until the new firm is driven out of the market, at which point the existing firm raises prices again. This pattern of pricing is aimed at deterring new firms from entering the market. However, in practice, it can be hard to figure out when pricing is predatory. Say that American Airlines is flying between two cities, and a new airline starts flying between the same two cities, at a lower price. If American Airlines cuts its price to match the new entrant, is this predatory pricing or is it just market competition at work? A commonly proposed rule is that if a firm is selling for less than its average variable cost—that is, at a price where it should be shutting down—then there is evidence for predatory pricing. However, calculating in the real world what costs are variable and what costs are fixed is often not obvious, either.

The Microsoft antitrust case embodies many of these gray areas in restrictive practices, as the next Clear It Up shows.

Clear It Up

Did Microsoft® engage in anticompetitive and restrictive practices?

The most famous restrictive practices case of recent years was a series of lawsuits by the U.S. government against Microsoft—lawsuits that some of Microsoft's competitors encouraged. All sides admitted that Microsoft's Windows program had a near-monopoly position in the market for the software used in general computer operating systems. All sides agreed that the software had many satisfied customers and that the computer software capabilities were compatible with Windows. Software that Microsoft and other companies produced had expanded dramatically in the 1990s. Having a **monopoly** or a near-monopoly is not necessarily illegal in and of itself, but in cases where one company controls a great deal of the market, antitrust regulators look at any allegations of restrictive practices with special care.

The antitrust regulators argued that Microsoft had gone beyond profiting from its software innovations and its dominant position in the software market for operating systems, and had tried to use its market power in operating systems software to take over other parts of the software industry. For example, the government argued that Microsoft had engaged in an anticompetitive form of exclusive dealing by threatening computer makers that, if they did not leave another firm's software off their machines (specifically, Netscape's Internet browser), then Microsoft would not sell them its operating system software. Government antitrust regulators accused Microsoft of tying together its Windows operating system software, where it had a monopoly, with its Internet Explorer browser software, where it did not have a monopoly, and thus using this bundling as an anticompetitive tool. The government also accused Microsoft of a form of predatory pricing; namely, giving away certain additional software products for free as part of Windows, as a way of driving out the competition from other software makers.

In April 2000, a federal court held that Microsoft's behavior had crossed the line into unfair competition, and recommended that the company be split into two competing firms. However, the court overturned that penalty on appeal, and in November 2002 Microsoft reached a settlement with the government that it would end its restrictive practices.

The concept of restrictive practices is continually evolving, as firms seek new ways to earn profits and government regulators define what is permissible. A situation where the law is evolving and changing is always somewhat troublesome, since laws are most useful and fair when firms know what they are in advance. In addition, since the law is open to interpretation, competitors who are losing out in the market can accuse successful firms of anticompetitive restrictive practices, and try to win through government regulation what they have failed to accomplish in the market.

Officials at the Federal Trade Commission and the Department of Justice are, of course, aware of these issues, but there is no easy way to resolve them.

11.3 | Regulating Natural Monopolies

By the end of this section, you will be able to:

- Evaluate the appropriate competition policy for a natural monopoly
- Interpret a graph of regulatory choices
- Contrast cost-plus and price cap regulation

Most true monopolies today in the U.S. are regulated, natural monopolies. A natural monopoly poses a difficult challenge for competition policy, because the structure of costs and demand makes competition unlikely or costly. A **natural monopoly** arises when average costs are declining over the range of production that satisfies market demand. This typically happens when fixed costs are large relative to variable costs. As a result, one firm is able to supply the total quantity demanded in the market at lower cost than two or more firms—so splitting up the natural monopoly would raise the average cost of production and force customers to pay more.

Public utilities, the companies that have traditionally provided water and electrical service across much of the United States, are leading examples of natural monopoly. It would make little sense to argue that a local water company should be divided into several competing companies, each with its own separate set of pipes and water supplies. Installing four or five identical sets of pipes under a city, one for each water company, so that each household could choose its own water provider, would be terribly costly. The same argument applies to the idea of having many competing companies for delivering electricity to homes, each with its own set of wires. Before the advent of wireless phones, the argument also applied to the idea of many different phone companies, each with its own set of phone wires running through the neighborhood.

The Choices in Regulating a Natural Monopoly

What then is the appropriate competition policy for a natural monopoly? **Figure 11.3** illustrates the case of natural monopoly, with a market demand curve that cuts through the downward-sloping portion of the **average cost curve**. Points A, B, C, and F illustrate four of the main choices for regulation. **Table 11.3** outlines the regulatory choices for dealing with a natural monopoly.

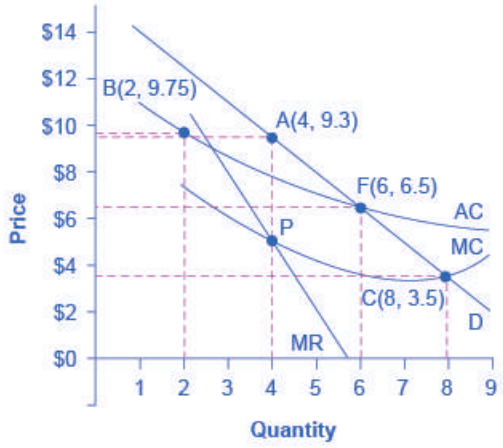


Figure 11.3 Regulatory Choices in Dealing with Natural Monopoly A natural monopoly will maximize profits by producing at the quantity where marginal revenue (MR) equals marginal costs (MC) and by then looking to the market demand curve to see what price to charge for this quantity. This monopoly will produce at point A, with a quantity of 4 and a price of 9.3. If antitrust regulators split this company exactly in half, then each half would produce at point B, with average costs of 9.75 and output of 2. The regulators might require the firm to produce where marginal cost crosses the market demand curve at point C. However, if the firm is required to produce at a quantity of 8 and sell at a price of 3.5, the firm will suffer from losses. The most likely choice is point F, where the firm is required to produce a quantity of 6 and charge a price of 6.5.

Quantity	Price	Total Revenue*	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
1	14.7	14.7	14.7	11.0	-	11.00
2	12.4	24.7	10.0	19.5	8.5	9.75
3	10.6	31.7	7.0	25.5	6.0	8.50
4	9.3	37.2	5.5	31.0	5.5	7.75
5	8.0	40.0	2.8	35.0	4.0	7.00
6	6.5	39.0	–1.0	39.0	4.0	6.50
7	5.0	35.0	–4.0	42.0	3.0	6.00
8	3.5	28.0	–7.0	45.5	3.5	5.70
9	2.0	18.0	–10.0	49.5	4.0	5.5

Table 11.3 Regulatory Choices in Dealing with Natural Monopoly (*We obtain total revenue by multiplying price and quantity. However, we have rounded some of the price values in this table for ease of presentation.)

The first possibility is to leave the natural monopoly alone. In this case, the monopoly will follow its normal approach to maximizing profits. It determines the quantity where MR = MC, which happens at point P at a quantity of 4. The firm then looks to point A on the demand curve to find that it can charge a price of 9.3 for that profit-maximizing quantity. Since the price is above the average cost curve, the natural monopoly would earn economic profits.

A second outcome arises if antitrust authorities decide to divide the company, so that the new firms can compete. As a simple example, imagine that the company is cut in half. Thus, instead of one large firm producing a quantity of 4, two half-size firms each produce a quantity of 2. Because of the declining average cost curve (AC), the average

cost of production for each of the half-size companies each producing 2, as point B shows, would be 9.75, while the average cost of production for a larger firm producing 4 would only be 7.75. Thus, the economy would become less productively efficient, since the good is produced at a higher average cost. In a situation with a downward-sloping average cost curve, two smaller firms will always have higher average costs of production than one larger firm for any quantity of total output. In addition, the antitrust authorities must worry that splitting the natural monopoly into pieces may be only the start of their problems. If one of the two firms grows larger than the other, it will have lower average costs and may be able to drive its competitor out of the market. Alternatively, two firms in a market may discover subtle ways of coordinating their behavior and keeping prices high. Either way, the result will not be the greater competition that was desired.

A third alternative is that regulators may decide to set prices and quantities produced for this industry. The regulators will try to choose a point along the market demand curve that benefits both consumers and the broader social interest. Point C illustrates one tempting choice: the regulator requires that the firm produce the quantity of output where marginal cost crosses the demand curve at an output of 8, and charge the price of 3.5, which is equal to **marginal cost** at that point. This rule is appealing because it requires price to be set equal to marginal cost, which is what would occur in a perfectly competitive market, and it would assure consumers a higher quantity and lower price than at the monopoly choice A. In fact, efficient allocation of resources would occur at point C, since the value to the consumers of the last unit bought and sold in this market is equal to the marginal cost of producing it.

Attempting to bring about point C through force of regulation, however, runs into a severe difficulty. At point C, with an output of 8, a price of 3.5 is below the average cost of production, which is 5.7, so if the firm charges a price of 3.5, it will be suffering losses. Unless the regulators or the government offer the firm an ongoing public subsidy (and there are numerous political problems with that option), the firm will lose money and go out of business.

Perhaps the most plausible option for the regulator is point F; that is, to set the price where AC crosses the demand curve at an output of 6 and a price of 6.5. This plan makes some sense at an intuitive level: let the natural monopoly charge enough to cover its average costs and earn a normal rate of profit, so that it can continue operating, but prevent the firm from raising prices and earning abnormally high monopoly profits, as it would at the monopoly choice A. Determining this level of output and price with the political pressures, time constraints, and limited information of the real world is much harder than identifying the point on a graph. For more on the problems that can arise from a centrally determined price, see the discussion of price floors and price ceilings in **Demand and Supply**.

Cost-Plus versus Price Cap Regulation

Regulators of public utilities for many decades followed the general approach of attempting to choose a point like F in **Figure 11.3**. They calculated the average cost of production for the water or electricity companies, added in an amount for the normal rate of profit the firm should expect to earn, and set the price for consumers accordingly. This method was known as **cost-plus regulation**.

Cost-plus regulation raises difficulties of its own. If producers receive reimbursement for their costs, plus a bit more, then at a minimum, producers have less reason to be concerned with high costs—because they can just pass them along in higher prices. Worse, firms under cost-plus regulation even have an incentive to generate high costs by building huge factories or employing many staff, because what they can charge is linked to the costs they incur.

Thus, in the 1980s and 1990s, some public utility regulators began to use **price cap regulation**, where the regulator sets a price that the firm can charge over the next few years. A common pattern was to require a price that declined slightly over time. If the firm can find ways of reducing its costs more quickly than the price caps, it can make a high level of profits. However, if the firm cannot keep up with the price caps or suffers bad luck in the market, it may suffer losses. A few years down the road, the regulators will then set a new series of price caps based on the firm's performance.

Price cap regulation requires delicacy. It will not work if the price regulators set the price cap unrealistically low. It may not work if the market changes dramatically so that the firm is doomed to incurring losses no matter what it does—say, if energy prices rise dramatically on world markets, then the company selling natural gas or heating oil to homes may not be able to meet price caps that seemed reasonable a year or two ago. However, if the regulators compare the prices with producers of the same good in other areas, they can, in effect, pressure a natural monopoly in one area to compete with the prices charged in other areas. Moreover, the possibility of earning greater profits or experiencing losses—instead of having an average rate of profit locked in every year by cost-plus regulation—can provide the natural monopoly with incentives for efficiency and innovation.

With natural monopoly, market competition is unlikely to take root, so if consumers are not to suffer the high prices

and restricted output of an unrestricted monopoly, government regulation will need to play a role. In attempting to design a system of price cap regulation with flexibility and incentive, government regulators do not have an easy task.

11.4 | The Great Deregulation Experiment

By the end of this section, you will be able to:

- Evaluate the effectiveness of price regulation and antitrust policy
- Explain regulatory capture and its significance

Governments at all levels across the United States have regulated prices in a wide range of industries. In some cases, like water and electricity that have natural monopoly characteristics, there is some room in economic theory for such regulation. However, once politicians are given a basis to intervene in markets and to choose prices and quantities, it is hard to know where to stop.

Doubts about Regulation of Prices and Quantities

Beginning in the 1970s, it became clear to policymakers of all political leanings that the existing price regulation was not working well. The United States carried out a great policy experiment—the **deregulation** that we discussed in **Monopoly**—removing government controls over prices and quantities produced in airlines, railroads, trucking, intercity bus travel, natural gas, and bank interest rates. The Clear It Up discusses the outcome of deregulation in one industry in particular—airlines.

What are the results of airline deregulation?

Why did the pendulum swing in favor of deregulation? Consider the airline industry. In the early days of air travel, no airline could make a profit just by flying passengers. Airlines needed something else to carry and the Postal Service provided that something with airmail. Thus, the first U.S. government regulation of the airline industry happened through the Postal Service, when in 1926 the Postmaster General began giving airlines permission to fly certain routes based on mail delivery needs—and the airlines took some passengers along for the ride. In 1934, the antitrust authorities charged the Postmaster General with colluding with the major airlines of that day to monopolize the nation's airways. In 1938, the U.S. government created the Civil Aeronautics Board (CAB) to regulate airfares and routes instead. For 40 years, from 1938 to 1978, the CAB approved all fares, controlled all entry and exit, and specified which airlines could fly which routes. There was zero entry of new airlines on the main routes across the country for 40 years, because the CAB did not think it was necessary.

In 1978, the Airline Deregulation Act took the government out of the business of determining airfares and schedules. The new law shook up the industry. Famous old airlines like Pan American, Eastern, and Braniff went bankrupt and disappeared. Some new airlines like People Express were created—and then vanished.

The greater competition from deregulation reduced airfares by about one-third over the next two decades, saving consumers billions of dollars a year. The average flight used to take off with just half its seats full; now it is two-thirds full, which is far more efficient. Airlines have also developed hub-and-spoke systems, where planes all fly into a central hub city at a certain time and then depart. As a result, one can fly between any of the spoke cities with just one connection—and there is greater service to more cities than before deregulation. With lower fares and more service, the number of air passengers doubled from the late 1970s to the start of the 2000s—an increase that, in turn, doubled the number of jobs in the airline industry. Meanwhile, with the watchful oversight of government safety inspectors, commercial air travel has continued to get safer over time.

The U.S. airline industry is far from perfect. For example, a string of mergers in recent years has raised concerns over how competition might be compromised.

One difficulty with government price regulation is what economists call **regulatory capture**, in which the firms that are supposedly regulated end up playing a large role in setting the regulations that they will follow. When the airline industry was regulated, for example, it suggested appointees to the regulatory board, sent lobbyists to argue with the board, provided most of the information on which the board made decisions, and offered well-paid jobs to at least some of the people leaving the board. In this situation, it is easy for regulators to poorly represent consumers. The result of regulatory capture is that government price regulation can often become a way for existing competitors to work together to reduce output, keep prices high, and limit competition.

The Effects of Deregulation

Deregulation, both of airlines and of other industries, has its negatives. The greater pressure of competition led to entry and exit. When firms went bankrupt or contracted substantially in size, they laid off workers who had to find other jobs. Market competition is, after all, a full-contact sport.

A number of major accounting scandals involving prominent corporations such as Enron, Tyco International, and WorldCom led to the **Sarbanes-Oxley Act** in 2002. The government designed Sarbanes-Oxley to increase confidence in financial information provided by public corporations to protect investors from accounting fraud.

The Great Recession, which began in late 2007, was caused at least in part by a global financial crisis, which began in the United States. The key component of the crisis was the creation and subsequent failure of several types of unregulated financial assets, such as collateralized mortgage obligations (CMOs, a type of mortgage-backed security), and credit default swaps (CDSs, insurance contracts on assets like CMOs that provided a payoff even if the holder of the CDS did not own the CMO). Private credit rating agencies such as Standard & Poors, Moody's, and Fitch rated many of these assets very safe.

The collapse of the markets for these assets precipitated the financial crisis and led to the failure of Lehman Brothers, a major investment bank, numerous large commercial banks, such as Wachovia, and even the Federal National Mortgage Corporation (Fannie Mae), which had to be nationalized—that is, taken over by the federal government. One response to the financial crisis was the **Dodd-Frank Act**, which majorly attempted to reform the financial system. The legislation's purpose, as noted on dodd-frank.com is:

> To promote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail," to protect the American taxpayer by ending bailouts, [and] to protect consumers from abusive financial services practices. . .

All market-based economies operate against a background of laws and regulations, including laws about enforcing contracts, collecting taxes, and protecting health and the environment. The government policies that we discussed in this chapter—like blocking certain anticompetitive mergers, ending restrictive practices, imposing price cap regulation on natural monopolies, and deregulation—demonstrate the role of government to strengthen the incentives that come with a greater degree of competition.

Bring it Home

More than Cooking, Heating, and Cooling

What did the Federal Trade Commission (FTC) decide on the Kinder Morgan / El Paso Corporation merger? After careful examination, federal officials decided there was only one area of significant overlap that might provide the merged firm with strong market power. The FTC approved the merger, provided Kinder Morgan divest itself of the overlap area. Tallgrass purchased Kinder Morgan Interstate Gas Transmission, Trailblazer Pipeline Co. LLC, two processing facilities in Wyoming, and Kinder Morgan's 50 percent interest in the Rockies Express Pipeline to meet the FTC requirements. The FTC was attempting to strike a balance between potential cost reductions resulting from economies of scale and concentration of market power.

Did the price of natural gas decrease? Yes, rather significantly. In 2010, the wellhead price of natural gas was $4.48 per thousand cubic foot. In 2012 the price had fallen to just $2.66. Was the merger responsible for the large drop in price? The answer is uncertain. The larger contributor to the sharp drop in price was the overall increase in the supply of natural gas. Increasingly, more natural gas was able to be recovered by fracturing shale deposits, a process called fracking. Fracking, which is controversial for environmental reasons, enabled

the recovery of known reserves of natural gas that previously were not economically feasible to tap. Kinder Morgan's control of 80,000-plus miles of pipeline likely made moving the gas from wellheads to end users smoother and allowed for an even greater benefit from the increased supply.

KEY TERMS

acquisition when one firm purchases another

antitrust laws laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones

bundling a situation in which multiple products are sold as one

concentration ratio an early tool to measure the degree of monopoly power in an industry; measures what share of the total sales in the industry are accounted for by the largest firms, typically the top four to eight firms

cost-plus regulation when regulators permit a regulated firm to cover its costs and to make a normal level of profit

exclusive dealing an agreement that a dealer will sell only products from one manufacturer

four-firm concentration ratio the percentage of the total sales in the industry that are accounted for by the largest four firms

Herfindahl-Hirschman Index (HHI) approach to measuring market concentration by adding the square of the market share of each firm in the industry

market share the percentage of total sales in the market

merger when two formerly separate firms combine to become a single firm

minimum resale price maintenance agreement an agreement that requires a dealer who buys from a manufacturer to sell for at least a certain minimum price

price cap regulation when the regulator sets a price that a firm cannot exceed over the next few years

regulatory capture when the supposedly regulated firms end up playing a large role in setting the regulations that they will follow and as a result, they "capture" the people usually through the promise of a job in that "regulated" industry once their term in government has ended

restrictive practices practices that reduce competition but that do not involve outright agreements between firms to raise prices or to reduce the quantity produced

tying sales a situation where a customer is allowed to buy one product only if the customer also buys another product

KEY CONCEPTS AND SUMMARY

11.1 Corporate Mergers

A corporate merger involves two private firms joining together. An acquisition refers to one firm buying another firm. In either case, two formerly independent firms become one firm. Antitrust laws seek to ensure active competition in markets, sometimes by preventing large firms from forming through mergers and acquisitions, sometimes by regulating business practices that might restrict competition, and sometimes by breaking up large firms into smaller competitors.

A four-firm concentration ratio is one way of measuring the extent of competition in a market. We calculate it by adding the market shares—that is, the percentage of total sales—of the four largest firms in the market. A Herfindahl-Hirschman Index (HHI) is another way of measuring the extent of competition in a market. We calculate it by taking the market shares of all firms in the market, squaring them, and then summing the total.

The forces of globalization and new communications and information technology have increased the level of competition that many firms face by increasing the amount of competition from other regions and countries.

11.2 Regulating Anticompetitive Behavior

Antitrust firms block authorities from openly colluding to form a cartel that will reduce output and raise prices. Companies sometimes attempt to find other ways around these restrictions and, consequently, many antitrust cases involve restrictive practices that can reduce competition in certain circumstances, like tie-in sales, bundling, and predatory pricing.

11.3 Regulating Natural Monopolies

In the case of a natural monopoly, market competition will not work well and so, rather than allowing an unregulated monopoly to raise price and reduce output, the government may wish to regulate price and/or output. Common examples of regulation are public utilities, the regulated firms that often provide electricity and water service.

Cost-plus regulation refers to government regulating a firm which sets the price that a firm can charge over a period of time by looking at the firm's accounting costs and then adding a normal rate of profit. Price cap regulation refers to government regulation of a firm where the government sets a price level several years in advance. In this case, the firm can either earn high profits if it manages to produce at lower costs or sell a higher quantity than expected or suffer low profits or losses if costs are high or it sells less than expected.

11.4 The Great Deregulation Experiment

The U.S. economy experienced a wave of deregulation in the late 1970s and early 1980s, when the government eliminated a number of regulations that had set prices and quantities produced in a number of industries. Major accounting scandals in the early 2000s and, more recently, the Great Recession have spurred new regulation to prevent similar occurrences in the future. Regulatory capture occurs when the regulated industries end up having a strong influence over what regulations exist.

SELF-CHECK QUESTIONS

1. Is it true that a merger between two firms that are not already in the top four by size can affect both the four-firm concentration ratio and the Herfindahl-Hirshman Index? Explain briefly.

2. Is it true that the four-firm concentration ratio puts more emphasis on one or two very large firms, while the Herfindahl-Hirshman Index puts more emphasis on all the firms in the entire market? Explain briefly.

3. Some years ago, two intercity bus companies, Greyhound Lines, Inc. and Trailways Transportation System, wanted to merge. One possible definition of the market in this case was "the market for intercity bus service." Another possible definition was "the market for intercity transportation, including personal cars, car rentals, passenger trains, and commuter air flights." Which definition do you think the bus companies preferred, and why?

4. As a result of globalization and new information and communications technology, would you expect that the definitions of markets that antitrust authorities use will become broader or narrower?

5. Why would a firm choose to use one or more of the anticompetitive practices described in **Regulating Anticompetitive Behavior**?

6. Urban transit systems, especially those with rail systems, typically experience significant economies of scale in operation. Consider the transit system data in **Table 11.4**. Note that the quantity is in millions of riders.

Demand:	Quantity	1	2	3	4	5	6	7	8	9	10
	Price	10	9	8	7	6	5	4	3	2	1
	Marginal Revenue	10	8	6	4	2	0	–2	–4	–6	–8
Costs:	Marginal Cost	9	6	5	3	2	3	4	5	7	10
	Average Cost	9	7.5	6.7	5.8	5	4.7	4.6	4.6	4.9	5.4

Table 11.4

Draw the demand, marginal revenue, marginal cost, and average cost curves. Do they have the normal shapes?

7. From the graph you drew to answer **Exercise 11.6**, would you say this transit system is a natural monopoly? Justify.

Use the following information to answer the next three questions. In the years before wireless phones, when telephone technology required having a wire running to every home, it seemed plausible that telephone service had diminishing average costs and might require regulation like a natural monopoly. For most of the twentieth century, the national U.S. phone company was AT&T, and the company functioned as a regulated monopoly. Think about the deregulation of the U.S. telecommunications industry that has occurred over the last few decades. (This is not a research assignment, but a thought assignment based on what you have learned in this chapter.)

8. What real world changes made the deregulation possible?

9. What are some of the benefits of the deregulation?

10. What might some of the negatives of deregulation be?

REVIEW QUESTIONS

11. What is a corporate merger? What is an acquisition?

12. What is the goal of antitrust policies?

13. How do we measure a four-firm concentration ratio? What does a high measure mean about the extent of competition?

14. How do we measure a Herfindahl-Hirshman Index? What does a low measure mean about the extent of competition?

15. Why can it be difficult to decide what a “market” is for purposes of measuring competition?

16. What is a minimum resale price maintenance agreement? How might it reduce competition and when might it be acceptable?

17. What is exclusive dealing? How might it reduce competition and when might it be acceptable?

18. What is a tie-in sale? How might it reduce competition and when might it be acceptable?

19. What is predatory pricing? How might it reduce competition, and why might it be difficult to tell when it should be illegal?

20. If public utilities are a natural monopoly, what would be the danger in deregulating them?

21. If public utilities are a natural monopoly, what would be the danger in splitting them into a number of separate competing firms?

22. What is cost-plus regulation?

23. What is price cap regulation?

24. What is deregulation? Name some industries that have been deregulated in the United States.

25. What is regulatory capture?

26. Why does regulatory capture reduce the persuasiveness of the case for regulating industries for the benefit of consumers?

CRITICAL THINKING QUESTIONS

27. Does either the four-firm concentration ratio or the HHI directly measure the amount of competition in an industry? Why or why not?

28. What would be evidence of serious competition between firms in an industry? Can you identify two highly competitive industries?

29. Can you think of any examples of successful predatory pricing in the real world?

30. If you were developing a product (like a web browser) for a market with significant barriers to entry, how would you try to get your product into the market successfully?

31. In the middle of the twentieth century, major U.S. cities had multiple competing city bus companies. Today, there is usually only one and it runs as a subsidized, regulated monopoly. What do you suppose caused the change?

32. Why are urban areas willing to subsidize urban transit systems? Does the argument for subsidies make sense to you?

33. Deregulation, like all changes in government policy, always has pluses and minuses. What do you think some of the minuses might be for airline deregulation?

34. Do you think it is possible for government to outlaw everything that businesses could do wrong? If so, why does government not do that? If not, how can regulation stay ahead of rogue businesses that push the limits of the system until it breaks?

PROBLEMS

35. Use **Table 11.5** to calculate the four-firm concentration ratio for the U.S. auto market. Does this indicate a concentrated market or not?

GM	19%
Ford	17%
Toyota	14%
Chrysler	11%

Table 11.5 Global Auto Manufacturers with Top Four U.S. Market Share, June 2013 (Source: http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013)

36. Use **Table 11.5** and **Table 11.6** to calculate the Herfindal-Hirschman Index for the U.S. auto market. Would the FTC approve a merger between GM and Ford?

Honda	10%
Nissan	7%
Hyundai	5%
Kia	4%
Subaru	3%
Volkswagen	3%

Table 11.6 Global Auto Manufacturers with additional U.S. Market Share, June 2013 (Source: http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013)

Use **Table 11.4** to answer the following questions.

37. If the transit system were allowed to operate as an unregulated monopoly, what output would it supply and what price would it charge?

38. If the transit system were regulated to operate with no subsidy (i.e., at zero economic profit), what approximate output would it supply and what approximate price would it charge?

39. If the transit system were regulated to provide the most allocatively efficient quantity of output, what output would it supply and what price would it charge? What subsidy would be necessary to insure this efficient provision of transit services?