



openstax™

Principles of Eco- nomics 3e.

Positive Externalities and Public Goods

13



FIGURE 13.1 View from Voyager I (<http://openstax.org//voyager1>) Launched by NASA on September 5, 1977, Voyager I's primary mission was to provide detailed images of Jupiter, Saturn, and their moons. It took this photograph of Jupiter on its journey. In August of 2012, Voyager I entered interstellar space—the first human-made object to do so—and it is expected to send data and images back to earth until 2025. Such a technological feat entails many economic principles. (Credit: modification of "Voyager's View of Jupiter's Great Red Spot" by NASA/JPL, Public Domain)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Why the Private Sector Underinvests in Technologies
- How Governments Can Encourage Innovation
- Public Goods

Introduction to Positive Externalities and Public Goods



BRING IT HOME

The Benefits of Voyager I Endure

The rapid growth of technology has increased our ability to access and process data, to navigate through a busy city, and to communicate with friends on the other side of the globe. The research and development efforts of citizens, scientists, firms, universities, and governments have truly revolutionized the modern economy. To get a sense of how far we have come in a short period of time, let's compare one of humankind's greatest achievements to the smartphone.

In 1977 the United States launched Voyager I, a spacecraft originally intended to reach Jupiter and Saturn, to send back photographs and other cosmic measurements. Voyager I, however, kept going, and going—past Jupiter and Saturn—right out of our solar system. At the time of its launch, Voyager had some of the most sophisticated computing processing power NASA could engineer (8,000 instructions per second), but today, we Earthlings use handheld devices that can process 14 billion instructions per second.

Still, the technology of today is a spillover product of the incredible feats NASA accomplished over forty years ago. NASA research, for instance, is responsible for the kidney dialysis and mammogram machines that we use today. Research in new technologies not only produces private benefits to the investing firm, or in this case to NASA, but it also creates benefits for the broader society. In this way, new knowledge often becomes what economists refer to as a public good. This leads us to the topic of this chapter—technology, positive externalities, public goods, and the role of government in encouraging innovation and the social benefits that it provides.

As economist Mariana Mazzucato explores in her well-known work *The Entrepreneurial State*, what makes a smartphone smart? What allows its apps to help you navigate new towns while getting updates about your home, all while your hands are on the steering wheel and your children are in the back seat watching their shows? For starters, the internet, cell tower networks, GPS, and voice activation. Each of these, and many other technologies we rely on, were developed with intensive government support. For example, GPS, which enables many cell phone functions beyond the frequently used mapping and ride-sharing applications, was developed by the U.S. Department of Defense over several generations of satellite tracking and complex computer algorithm development. The U.S. government still provides GPS for many of the world's users.

We do not often think of the government when we consider our leading products and entrepreneurs. We think of Apple, Google, Lyft, Tesla, Fitbit, and so on—creative innovators who built on the tools provided by these government efforts, using them in transformative ways. We may not think of the estimated \$19 billion per year that the U.S. spends to maintain the GPS system, but we would certainly think of it if it suddenly went away. (Beyond the impact on our daily lives, economists estimate U.S. businesses alone would lose about \$1 billion per day without GPS.)

Mazzucato is one of several prominent economists advocating for an embrace of continued government-sponsored innovations in order to build economic prosperity, reduce inequality, and manage ongoing challenges such as drought, coastal changes, and extreme weather. She argues that competitive, private sector markets are often resistant to the risks involved with large-scale innovation, because failed experiments and lack of uptake lead to massive corporate and personal losses. Governments can take on riskier research and development projects. Because government spending is fueled by taxpayers, and all innovation leads to some level of employment change, these proposals are certainly complex and challenging to implement.

This chapter deals with some of these issues: Will private companies be willing to invest in new technology? In what ways does new technology have positive externalities? What motivates inventors? What role should government play in encouraging research and technology? Are there certain types of goods that markets fail to provide efficiently, and that only government can produce? What happens when consumption or production of a product creates positive externalities? Why is it unsurprising when we overuse a common resource, like marine fisheries?

13.1 Investments in Innovation

LEARNING OBJECTIVES

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

- Identify the positive externalities of new technology.
- Explain the difference between private benefits and social benefits and give examples of each.
- Calculate and analyze rates of return

Market competition can provide an incentive for discovering new technology because a firm can earn higher

profits by finding a way to produce products more cheaply or to create products with characteristics consumers want. As Gregory Lee, CEO of Samsung said, “Relentless pursuit of new innovation is the key principle of our business and enables consumers to discover a world of possibilities with technology.” An innovative firm knows that it will usually have a temporary edge over its competitors and thus an ability to earn above-normal profits before competitors can catch up.

In certain cases, however, competition can discourage new technology, especially when other firms can quickly copy a new idea. Consider a pharmaceutical firm deciding to develop a new drug. On average, it can cost \$800 million and take more than a decade to discover a new drug, perform the necessary safety tests, and bring the drug to market. If the research and development (R&D) effort fails—and every R&D project has some chance of failure—then the firm will suffer losses and could even be driven out of business. If the project succeeds, then the firm’s competitors may figure out ways of adapting and copying the underlying idea, but without having to pay the costs themselves. As a result, the innovative company will bear the much higher costs of the R&D and will enjoy at best only a small, temporary advantage over the competition.

Many inventors over the years have discovered that their inventions brought them less profit than they might have reasonably expected.

- Eli Whitney (1765–1825) invented the cotton gin, but then southern cotton planters built their own seed-separating devices with a few minor changes in Whitney’s design. When Whitney sued, he found that the courts in southern states would not uphold his patent rights.
- Thomas Edison (1847–1931) still holds the record for most patents granted to an individual. His first invention was an automatic vote counter, and despite the social benefits, he could not find a government that wanted to buy it.
- Gordon Gould came up with the idea behind the laser in 1957. He put off applying for a patent and, by the time he did apply, other scientists had laser inventions of their own. A lengthy legal battle resulted, in which Gould spent \$100,000 on lawyers, before he eventually received a patent for the laser in 1977. Compared to the enormous social benefits of the laser, Gould received relatively little financial reward.
- In 1936, Alan Turing delivered a paper titled, “On Computable Numbers, with an Application to the Entscheidungsproblem,” in which he presented the notion of a universal machine (later called the “Universal Turing Machine,” and then the “Turing machine”) capable of computing anything that is computable. The central concept of the modern computer was based on Turing’s paper. Today scholars widely consider Turing as the father of theoretical computer science and artificial intelligence; however, the UK government prosecuted Turing in 1952 for engaging in same-sex sexual acts and gave him the choice of chemical castration or prison. Turing chose castration and died in 1954 from cyanide poisoning.

A variety of studies by economists have found that the original inventor receives one-third to one-half of the total economic benefits from innovations, while other businesses and new product users receive the rest.

The Positive Externalities of New Technology

Will private firms in a market economy underinvest in research and technology? If a firm builds a factory or buys a piece of equipment, the firm receives all the economic benefits that result from the investments. However, when a firm invests in new technology, the **private benefits**, or profits, that the firm receives are only a portion of the overall social benefits. The **social benefits** of an innovation account for the value of all the positive externalities of the new idea or product, whether enjoyed by other companies or society as a whole, as well as the private benefits the firm that developed the new technology receives. As you learned in [Environmental Protection and Negative Externalities](#), **positive externalities** are beneficial spillovers to a third party, or parties.

Consider the example of the Big Drug Company, which is planning its R&D budget for the next year. Economists and scientists working for Big Drug have compiled a list of potential research and development projects and estimated rates of return. (The rate of return is the estimated payoff from the project.) [Figure 13.2](#) shows how

the calculations work. The downward-sloping D_{Private} curve represents the firm's demand for financial capital and reflects the company's willingness to borrow to finance research and development projects at various interest rates. Suppose that this firm's investment in research and development creates a spillover benefit to other firms and households. After all, new innovations often spark other creative endeavors that society also values. If we add the spillover benefits society enjoys to the firm's private demand for financial capital, we can draw D_{Social} that lies above D_{Private} .

If there were a way for the firm to fully monopolize those social benefits by somehow making them unavailable to the rest of us, the firm's private demand curve would be the same as society's demand curve. According to [Figure 13.2](#) and [Table 13.1](#), if the going rate of interest on borrowing is 8%, and the company can receive the private benefits of innovation only, then the company would finance \$30 million. Society, at the same rate of 8%, would find it optimal to have \$52 million of borrowing. Unless there is a way for the company to fully enjoy the total benefits, then it will borrow less than the socially optimal level of \$52 million.

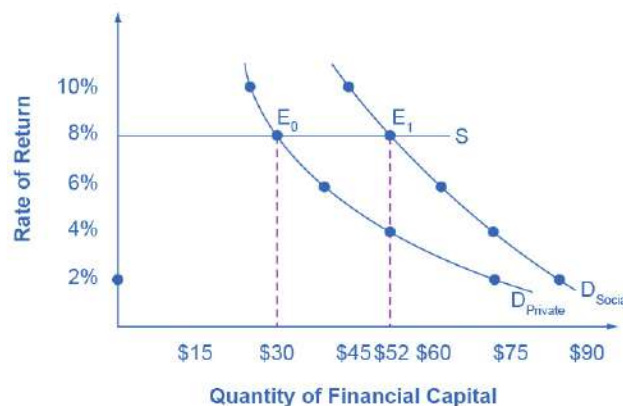


FIGURE 13.2 Positive Externalities and Technology Big Drug faces a cost of borrowing of 8%. If the firm receives only the private benefits of investing in R&D, then we show its demand curve for financial capital by D_{Private} , and the equilibrium will occur at \$30 million. Because there are spillover benefits, society would find it optimal to have \$52 million of investment. If the firm could keep the social benefits of its investment for itself, its demand curve for financial capital would be D_{Social} and it would be willing to borrow \$52 million.

Rate of Return	D_{Private} (in millions)	D_{Social} (in millions)
2%	\$72	\$84
4%	\$52	\$72
6%	\$38	\$62
8%	\$30	\$52
10%	\$26	\$44

TABLE 13.1 Return and Demand for Capital

Big Drug's original demand for financial capital (D_{Private}) is based on the profits the firm receives. However, other pharmaceutical firms and health care companies may learn new lessons about how to treat certain medical conditions and are then able to create their own competing products. The social benefit of the drug takes into account the value of all the drug's positive externalities. If Big Drug were able to gain this social return instead of other companies, its demand for financial capital would shift to the demand curve D_{Social} , and it would be willing to borrow and invest \$52 million. However, if Big Drug is receiving only 50 cents of each

dollar of social benefits, the firm will not spend as much on creating new products. The amount it would be willing to spend would fall somewhere in between D_{Private} and D_{Social} .

Why Invest in Human Capital?

The investment in anything, whether it is the construction of a new power plant or research in a new cancer treatment, usually requires a certain upfront cost with an uncertain future benefit. The investment in education, or human capital, is no different. Over the span of many years, a student and her family invest significant amounts of time and money into that student's education. The idea is that higher levels of educational attainment will eventually serve to increase the student's future productivity and subsequent ability to earn. Once the student crunches the numbers, does this investment pay off for her?

Almost universally, economists have found that the answer to this question is a clear "Yes." For example, several studies of the return to education in the United States estimate that the rate of return to a college education is approximately 10-15%. Data in [Table 13.2](#), from the U.S. Bureau of Labor Statistics' *Usual Weekly Earnings of Wage and Salary Workers, Fourth Quarter 2021*, demonstrate that median weekly earnings are higher for workers who have completed more education. While these rates of return will beat equivalent investments in Treasury bonds or savings accounts, the estimated returns to education go primarily to the individual worker, so these returns are **private rates of return** to education.

	Less than a High School Degree	High School Degree, No College	Bachelor's Degree or Higher
Median Weekly Earnings (full-time workers over the age of 25)	\$651	\$831	\$1,467

TABLE 13.2 Usual Weekly Earnings of Wage and Salary Workers, Fourth Quarter 2021 (Source: <https://www.bls.gov/news.release/pdf/wkyeng.pdf>)

What does society gain from investing in the education of another student? After all, if the government is spending taxpayer dollars to subsidize public education, society should expect some kind of return on that spending. Economists like George Psacharopoulos have found that, across a variety of nations, the **social rate of return** on schooling is also positive. After all, positive externalities exist from investment in education. While not always easy to measure, according to Walter McMahon, the positive externalities to education typically include better health outcomes for the population, lower levels of crime, a cleaner environment and a more stable, democratic government. For these reasons, many nations have chosen to use taxpayer dollars to subsidize primary, secondary, and higher education. Education clearly benefits the person who receives it, but a society where most people have a good level of education provides positive externalities for all.

Other Examples of Positive Externalities

Although technology may be the most prominent example of a positive externality, it is not the only one. For example, vaccinations against disease are not only a protection for the individual, but they have the positive spillover of protecting others who may become infected. When a number of homes in a neighborhood are modernized, updated, and restored, not only does it increase the homes' value, but other property values in the neighborhood may increase as well.

The appropriate public policy response to a positive externality, like a new technology, is to help the party creating the positive externality receive a greater share of the social benefits. In the case of vaccines, like flu shots, an effective policy might be to provide a subsidy to those who choose to get vaccinated.

[Figure 13.3](#) shows the market for flu shots. The market demand curve D_{Market} for flu shots reflects only the marginal private benefits (MPB) that the vaccinated individuals receive from the shots. Assuming that there are no spillover costs in the production of flu shots, the market supply curve is given by the marginal private

cost (MPC) of producing the vaccinations.

The equilibrium quantity of flu shots produced in the market, where MPB is equal to MPC, is Q_{Market} and the price of flu shots is P_{Market} . However, spillover benefits exist in this market because others, those who chose not to purchase a flu shot, receive a positive externality in the form of a reduced chance of contracting the flu. When we add the spillover benefits to the marginal private benefit of flu shots, the marginal social benefit (MSB) of flu shots is given by D_{Social} . Because the MSB is greater than MPB, we see that the socially optimal level of flu shots is greater than the market quantity (Q_{Social} exceeds Q_{Market}) and the corresponding price of flu shots, if the market were to produce Q_{Social} , would be at P_{Social} . Unfortunately, the marketplace does not recognize the positive externality and flu shots will go under-produced and under-consumed.

How can government try to move the market level of output closer to the socially desirable level of output? One policy would be to provide a subsidy, like a voucher, to any citizen who wishes to get vaccinated. This voucher would act as “income” that one could use to purchase only a flu shot and, if the voucher were exactly equal to the per-unit spillover benefits, would increase market equilibrium to a quantity of Q_{Social} and a price of P_{Social} where MSB equals MSC (which equals MPC given the assumption that there are no spillover costs in producing the vaccine). Suppliers of the flu shots would receive payment of P_{Social} per vaccination, while consumers of flu shots would redeem the voucher and only pay a price of P_{Subsidy} . When the government uses a subsidy in this way, it produces the socially optimal quantity of vaccinations.

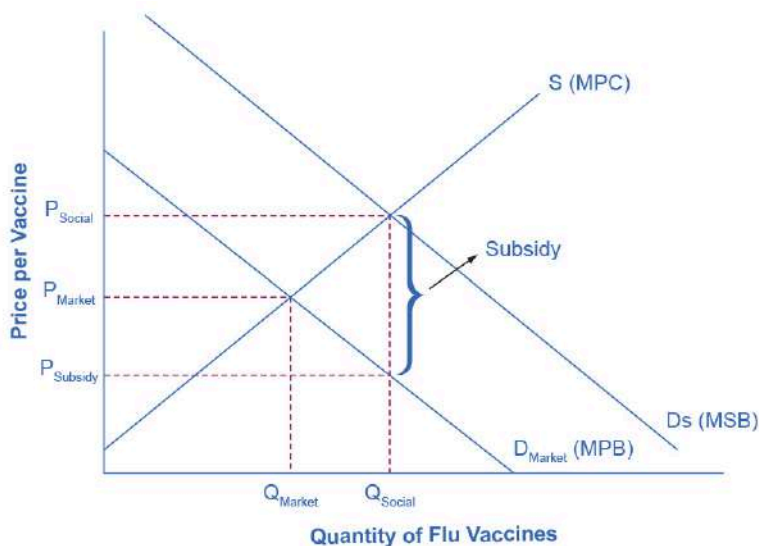


FIGURE 13.3 The Market for Flu Shots with Spillover Benefits (A Positive Externality) The market demand curve does not reflect the positive externality of flu vaccinations, so only Q_{Market} will be exchanged. This outcome is inefficient because the marginal social benefit exceeds the marginal social cost. If the government provides a subsidy to consumers of flu shots, equal to the marginal social benefit minus the marginal private benefit, the level of vaccinations can increase to the socially optimal quantity of Q_{Social} .

Societal Change as an Innovation Outcome

Economist Carlota Perez draws on the lessons of past innovations to understand the current state of our economy. She demonstrates that prior technological turning points, such as the proliferation of railroads and the emergence of mass production, created initial periods of employment and wealth shifting but eventually led to greater well-being and economic growth. After difficult transition periods and sometimes economic meltdowns during the “installment” phase of widespread new technologies, many economies and the people within them have benefited from prolonged periods of economic and lifestyle improvement, including lower unemployment and better quality of life.

Most prior innovation periods, such as the Industrial Revolution, had one significant downside: negative

impacts on the environment, such as pollution and habitat destruction. Perez notes that our current revolution—in information and communications technology (ICT)—has the potential for significant positive externalities related to the environment. ICT is shifting many areas of society (and therefore industry) to digital experiences and services that do not require fossil fuels or similar natural resources. Vehicle sharing, product rental-reuse networks, and new manufacturing methods offer the promise of far less consumable consumption. And even though the appearance of delivery trucks and shipping boxes gives the impression of environmental damage, most studies indicate that online shopping is better for the environment than individuals shopping in person. (This is partly attributed to greater efficiency in a few trucks driving to a neighborhood rather than everyone in the neighborhood driving to several stores.) Consumers and governments can spur on those environmental benefits by choosing or partnering with companies that focus on furthering their environmental impact, such as by using solar power to fuel their computer servers or by using electrically powered delivery trucks.

Like other innovations, ICT has created some employment and economic opportunities while it has reduced others. Increased globalization and efficiencies have shuttered businesses and reduced wages in some areas. Perez's research indicates that those types of employment shifts can be managed through proper regulation and investment (especially in human capital), particularly as firms in the relevant industries become mature and profitable. The prospects aren't simple: ICT has created megafirms like Amazon and Apple, which despite pleasing their consumers can wield significant power over governments and employees. But on the environmental and societal front at least, ICT has offered a wealth of opportunities and externalities.

13.2 How Governments Can Encourage Innovation

LEARNING OBJECTIVES

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

- Explain the effects of intellectual property rights on social and private rates of return.
- Identify three U.S. Government policies and explain how they encourage innovation

A number of different government policies can increase the incentives to innovate, including: guaranteeing intellectual property rights, government assistance with the costs of research and development, and cooperative research ventures between universities and companies.

Intellectual Property Rights

One way to increase new technology is to guarantee the innovator an exclusive right to that new product or process. **Intellectual property** rights include patents, which give the inventor the exclusive legal right to make, use, or sell the invention for a limited time, and copyright laws, which give the author an exclusive legal right over works of literature, music, film/video, and pictures. For example, if a pharmaceutical firm has a patent on a new drug, then no other firm can manufacture or sell that drug for 20 years, unless the firm with the patent grants permission. Without a patent, the pharmaceutical firm would have to face competition for any successful products, and could earn no more than a normal rate of profit. With a patent, a firm is able to earn monopoly profits on its product for a period of time—which offers an incentive for research and development. In general, how long can “a period of time” be? The Clear It Up discusses patent and copyright protection timeframes for some works you might know.



CLEAR IT UP

How long is Mickey Mouse protected from being copied?

All patents and copyrights are scheduled to end someday. In 2003, copyright protection for Mickey Mouse was scheduled to run out. Once the copyright had expired, anyone would be able to copy Mickey Mouse cartoons or draw and sell new ones. In 1998, however, Congress passed the Sonny Bono Copyright Term Extension Act. For copyrights owned by companies or other entities, it increased or extended the copyright from 75 years to 95 years

after publication. For copyrights owned by individuals, it increased or extended the copyright coverage from 50 years to 70 years after death. Along with protecting Mickey for another 20 years, the copyright extension affected about 400,000 books, movies, and songs.

Figure 13.4 illustrates how the total number of patent applications filed with the U.S. Patent and Trademark Office, as well as the total number of patents granted, surged in the mid-1990s with the invention of the internet, and is still going strong today.

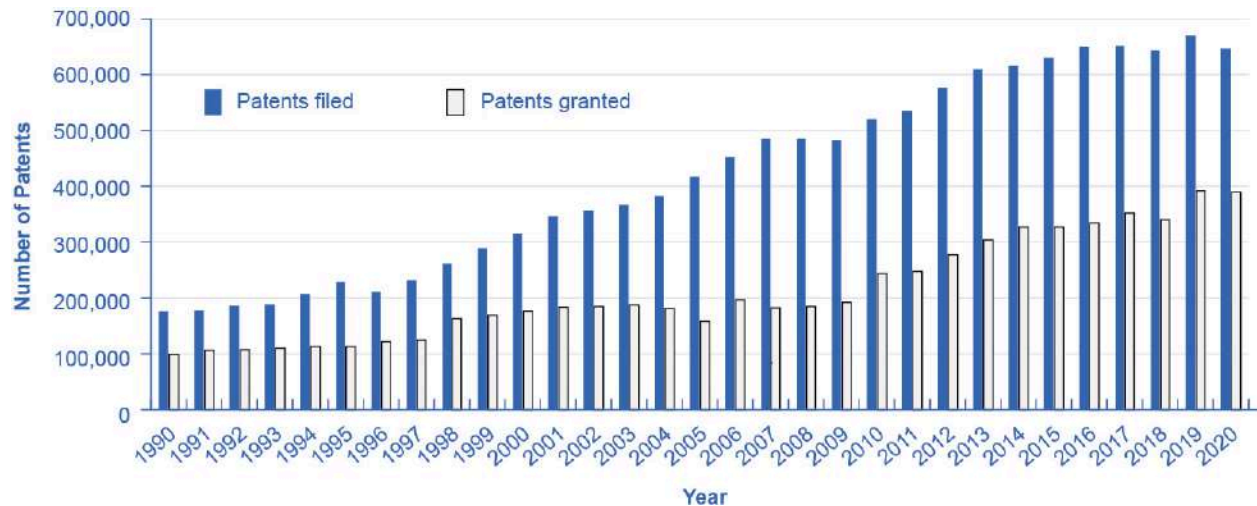


FIGURE 13.4 Patents Filed and Granted, 1981–2012 The number of applications filed for patents increased substantially beginning in the 1990s, due in part to the invention of the internet, which has led to many other inventions and to the 1998 Copyright Term Extension Act. (Source: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm)

While patents provide an incentive to innovate by protecting the innovator, they are not perfect. For example:

- In countries that already have patents, economic studies show that inventors receive only one-third to one-half of the total economic value of their inventions.
- In a fast-moving high-technology industry like biotechnology or semiconductor design, patents may be almost irrelevant because technology is advancing so quickly.
- Not every new idea can be protected with a patent or a copyright—for example, a new way of organizing a factory or a new way of training employees.
- Patents may sometimes cover too much or be granted too easily. In the early 1970s, Xerox had received over 1,700 patents on various elements of the photocopy machine. Every time Xerox improved the photocopier, it received a patent on the improvement.
- The 20-year time period for a patent is somewhat arbitrary. Ideally, a patent should cover a long enough period of time for the inventor to earn a good return, but not so long that it allows the inventor to charge a monopoly price permanently.

Because patents are imperfect and do not apply well to all situations, alternative methods of improving the rate of return for inventors of new technology are desirable. The following sections describe some of these possible alternative policies.

Policy #1: Government Spending on Research and Development

If the private sector does not have sufficient incentive to carry out research and development, one possibility is for the government to fund such work directly. Government spending can provide direct financial support for research and development (R&D) conducted at colleges and universities, nonprofit research entities, and sometimes by private firms, as well as at government-run laboratories. While government spending on

research and development produces technology that is broadly available for firms to use, it costs taxpayers money and can sometimes be directed more for political than for scientific or economic reasons.

LINK IT UP

Visit the NASA [website \(http://openstax.org/l/NASA\)](http://openstax.org/l/NASA) and the USDA [website \(http://openstax.org/l/USDA\)](http://openstax.org/l/USDA) to read about government research that would not take place were it left to firms, due to the externalities.

The first column of [Table 13.3](#) shows the sources of total U.S. spending on research and development. The second column shows the total dollars of R&D funding by each source. The third column shows that, relative to the total amount of funding, 22.7% comes from the federal government, about 69% of R&D is done by industry, and less than 4% is done by universities and colleges. (The percentages below do not add up to exactly 100% due to rounding.)

Sources of R&D Funding	Amount (\$ billions)	Percent of the Total
Federal government	\$129.6	21.4%
Industry	\$426.0	70.3%
Universities and colleges	\$20.7	3.4%
Nonprofits	\$25.0	4.1%
Nonfederal government	\$4.8	0.8%
<i>Total</i>	<i>\$606.1</i>	

TABLE 13.3 U.S. Research and Development Expenditures, 2018

(Source: <https://nces.nsf.gov/pubs/nsf21324>)

In the 1960s the federal government paid for about two-thirds of the nation's R&D. Over time, the U.S. economy has come to rely much more heavily on industry-funded R&D. The federal government has tried to focus its direct R&D spending on areas where private firms are not as active. One difficulty with direct government support of R&D is that it inevitably involves political decisions about which projects are worthy. The scientific question of whether research is worthwhile can easily become entangled with considerations like the location of the congressional district in which the research funding is spent.

Policy #2: Tax Breaks for Research and Development

A complementary approach to supporting R&D that does not involve the government's close scrutiny of specific projects is to give firms a reduction in taxes depending on how much research and development they do. The federal government refers to this policy as the research and experimentation (R&E) tax credit. According to the Treasury Department: "... the R&E Credit is also a cost-effective policy for stimulating additional private sector investment. Most recent studies find that each dollar of foregone tax revenue through the R&E Tax Credit causes firms to invest at least a dollar in R&D, with some studies finding a benefit to cost ratio of 2 or 2.96."

LINK IT UP

Visit this [website \(http://openstax.org/l/REtaxcredit\)](http://openstax.org/l/REtaxcredit) for more information on how the R&E Tax Credit encourages investment.

Policy #3 Cooperative Research

State and federal governments support research in a variety of ways. For example, United for Medical Research, a coalition of groups that seek funding for the National Institutes of Health, (which is supported by federal grants), states: “NIH-supported research added \$69 billion to our GDP and supported seven million jobs in 2011 alone.” The United States remains the leading sponsor of medical-related research, spending \$117 billion in 2011. Other institutions, such as the National Academy of Sciences and the National Academy of Engineering, receive federal grants for innovative projects. The Agriculture and Food Research Initiative (AFRI) at the United States Department of Agriculture awards federal grants to projects that apply the best science to the most important agricultural problems, from food safety to childhood obesity. Cooperation between government-funded universities, academies, and the private sector can spur product innovation and create whole new industries.

13.3 Public Goods

LEARNING OBJECTIVES

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

- Identify a public good using nonexcludable and non-rival as criteria
- Explain the free rider problem
- Identify several sources of public goods

Even though new technology creates positive externalities so that perhaps one-half or two-thirds of the social benefit of new inventions spills over to others, the inventor still receives some private return. What about a situation where the positive externalities are so extensive that private firms could not expect to receive any of the social benefit? We call this kind of good a **public good**. Spending on national defense is a good example of a public good. Let’s begin by defining the characteristics of a public good and discussing why these characteristics make it difficult for private firms to supply public goods. Then we will see how government may step in to address the issue.

The Definition of a Public Good

Economists have a strict definition of a public good, and it does not necessarily include all goods financed through taxes. To understand the defining characteristics of a public good, first consider an ordinary private good, like a piece of pizza. We can buy and sell a piece of pizza fairly easily because it is a separate and identifiable item. However, public goods are not separate and identifiable in this way.

Instead, public goods have two defining characteristics: they are nonexcludable and non-rival. The first characteristic, that a public good is **nonexcludable**, means that it is costly or impossible to exclude someone from using the good. If Larry buys a private good like a piece of pizza, then he can exclude others, like Lorna, from eating that pizza. However, if national defense is provided, then it includes everyone. Even if you strongly disagree with America’s defense policies or with the level of defense spending, the national defense still protects you. You cannot choose to be unprotected, and national defense cannot protect everyone else and exclude you.

The second main characteristic of a public good, that it is **non-rival**, means that when one person uses the public good, another can also use it. With a private good like pizza, if Max is eating the pizza then Michelle cannot also eat it; that is, the two people are rivals in consumption. With a public good like national defense, Max’s consumption of national defense does not reduce the amount left for Michelle, so they are non-rival in this area.

A number of government services are examples of public goods. For instance, it would not be easy to provide fire and police service so that some people in a neighborhood would be protected from the burning and burglary of their property, while others would not be protected at all. Protecting some necessarily means protecting others, too.

Positive externalities and public goods are closely related concepts. Public goods have positive externalities, like police protection or public health funding. Not all goods and services with positive externalities, however, are public goods. Investments in education have huge positive spillovers but can be provided by a private company. Private companies can invest in new inventions such as the Apple iPad and reap profits that may not capture all of the social benefits. We can also describe patents as an attempt to make new inventions into private goods, which are excludable and rivalrous, so that no one but the inventor can use them during the length of the patent.

The Free Rider Problem of Public Goods

Private companies find it difficult to produce public goods. If a good or service is nonexcludable, like national defense, so that it is impossible or very costly to exclude people from using this good or service, then how can a firm charge people for it?

LINK IT UP

Visit this [website \(http://openstax.org/l/freerider\)](http://openstax.org/l/freerider) to read about a connection between free riders and “bad music.”

When individuals make decisions about buying a public good, a **free rider** problem can arise, in which people have an incentive to let others pay for the public good and then to “free ride” on the purchases of others. We can express the free rider problem in terms of the prisoner’s dilemma game, which we discussed as a representation of oligopoly in [Monopolistic Competition and Oligopoly](#).

There is a dilemma with the Prisoner’s Dilemma, though. See the Work It Out feature.

WORK IT OUT

The Problem with the Prisoner’s Dilemma

Suppose two people, Rachel and Samuel, are considering purchasing a public good. The difficulty with the prisoner’s dilemma arises as each person thinks through their strategic choices.

Step 1. Rachel reasons in this way: If Samuel does not contribute, then I would be a fool to contribute. However, if Samuel does contribute, then I can come out ahead by not contributing.

Step 2. Either way, I should choose not to contribute, and instead hope that I can be a free rider who uses the public good paid for by Samuel.

Step 3. Samuel reasons the same way about Rachel.

Step 4. When both people reason in that way, the public good never gets built, and there is no movement to the option where everyone cooperates—which is actually best for all parties.

The Role of Government in Paying for Public Goods

The key insight in paying for public goods is to find a way of assuring that everyone will make a contribution and to prevent free riders. For example, if people come together through the political process and agree to pay taxes and make group decisions about the quantity of public goods, they can defeat the free rider problem by requiring, through the law, that everyone contributes.

However, government spending and taxes are not the only way to provide public goods. In some cases, markets can produce public goods. For example, think about radio. It is nonexcludable, since once the radio signal is broadcast, it would be very difficult to stop someone from receiving it. It is non-rival, since one person listening to the signal does not prevent others from listening as well. Because of these features, it is practically

impossible to charge listeners directly for listening to conventional radio broadcasts.

Radio has found a way to collect revenue by selling advertising, which is an indirect way of “charging” listeners by taking up some of their time. Ultimately, consumers who purchase the goods advertised are also paying for the radio service, since the station builds in the cost of advertising into the product cost. In a more recent development, satellite radio companies, such as SiriusXM, charge a regular subscription fee for streaming music without commercials. In this case, however, the product is excludable—only those who pay for the subscription will receive the broadcast.

Some public goods will also have a mixture of public provision at no charge along with fees for some purposes, like a public city park that is free to use, but the government charges a fee for parking your car, for reserving certain picnic grounds, and for food sold at a refreshment stand.

LINK IT UP

Read this [article \(http://openstax.org/l/governmentpay\)](http://openstax.org/l/governmentpay) to find out what economists say the government should pay for.

In other cases, we can use social pressures and personal appeals, rather than the force of law, to reduce the number of free riders and to collect resources for the public good. For example, neighbors sometimes form an association to carry out beautification projects or to patrol their area after dark to discourage crime. In low-income countries, where social pressure strongly encourages all farmers to participate, farmers in a region may come together to work on a large irrigation project that will benefit all. We can view many fundraising efforts, including raising money for local charities and for the endowments of colleges and universities, as an attempt to use social pressure to discourage free riding and to generate the outcome that will produce a public benefit.

Common Resources and the “Tragedy of the Commons”

There are some goods that do not fall neatly into the categories of private good or public good. While it is easy to classify a pizza as a private good and a city park as a public good, what about an item that is nonexcludable and rivalrous, such as the queen conch?

In the Caribbean, the queen conch is a large marine mollusk that lives in shallow waters of sea grass. These waters are so shallow, and so clear, that a single diver may harvest many conch in a single day. Not only is conch meat a local delicacy and an important part of the local diet, but artists use the large ornate shells and craftsmen transform them. Because almost anyone with a small boat, snorkel, and mask, can participate in the conch harvest, it is essentially nonexcludable. At the same time, fishing for conch is rivalrous. Once a diver catches one conch another diver cannot catch it.

We call goods that are nonexcludable and rivalrous common resources. Because the waters of the Caribbean are open to all conch fishermen, and because any conch that *you* catch is a conch that *I* cannot catch, fishermen tend to overharvest common resources like the conch.

The problem of overharvesting common resources is not a new one, but ecologist Garret Hardin put the tag “tragedy of the commons” to the problem in a 1968 article in the magazine *Science*. Economists view this as a problem of property rights. Since nobody owns the ocean, or the conch that crawl on the sand beneath it, no one individual has an incentive to protect that resource and responsibly harvest it. To address the issue of overharvesting conch and other marine fisheries, economists have advocated simple devices like fishing licenses, harvest limits, and shorter fishing seasons. One approach that has been turned to more recently is the implementation of catch shares, whereby regulators establish a total allowable catch, and then fishermen are allocated a portion of that total allowable catch. Catch shares appear to slow the race to fish. When the population of a species drops to critically low numbers, governments have even banned the harvest until biologists determine that the population has returned to sustainable levels. In fact, such is the case with the

conch, the harvesting of which the government has effectively banned in the United States since 1986.

The tragedy of the commons is a frequent economic and social framework for discussions about a range of common resources, even extending into digital resources such as open media repositories and online libraries. Prominent economist Elinor Ostrom, the first woman to receive the Nobel Prize in Economics, proposed an alternate version, sometimes referred to as the "non-tragedy of the commons." After extensive fieldwork in areas as diverse as Indonesia, Kenya, Maine (U.S.), and Nepal, she challenged the notion that people would only avoid depletion of common resources if they were forced to by regulatory laws and property rights. She noted that farmers working shared land could communicate and cooperate in order to maximize and preserve the fields over time. She argued that when those who benefit most from a resource are in close proximity to it (like a farm field that directly serves a town), the resource is better managed without external influence.

LINK IT UP

Visit this [website \(http://openstax.org/l/queenconch\)](http://openstax.org/l/queenconch) for more on the queen conch industry.

Positive Externalities in Public Health Programs

One of the most remarkable changes in the standard of living in the last several centuries is that people are living longer. Scientists believe that, thousands of years ago, human life expectancy ranged between 20 to 30 years. By 1900, average life expectancy in the United States was 47 years. By 2015, life expectancy was 79 years; due to COVID-19, life expectancy declined slightly to 77 years in 2020. Most of the gains in life expectancy in the history of the human race happened in the twentieth century.

The rise in life expectancy seems to stem from three primary factors. First, systems for providing clean water and disposing of human waste helped to prevent the transmission of many diseases. Second, changes in public behavior have advanced health. Early in the twentieth century, for example, people learned the importance of boiling bottles before using them for food storage and baby's milk, washing their hands, and protecting food from flies. More recent behavioral changes include reducing the number of people who smoke tobacco and precautions to limit sexually transmitted diseases. Third, medicine has played a large role. Scientists developed immunizations for diphtheria, cholera, pertussis, tuberculosis, tetanus, and yellow fever between 1890 and 1930. Penicillin, discovered in 1941, led to a series of other antibiotic drugs for bringing infectious diseases under control. In recent decades, drugs that reduce the risks of high blood pressure have had a dramatic effect in extending lives.

These advances in public health have all been closely linked to positive externalities and public goods. Public health officials taught hygienic practices to mothers in the early 1900s and encouraged less smoking in the late 1900s. Government funded many public sanitation systems and storm sewers because they have the key traits of public goods. In the twentieth century, many medical discoveries emerged from government or university-funded research. Patents and intellectual property rights provided an additional incentive for private inventors. The reason for requiring immunizations, phrased in economic terms, is that it prevents spillovers of illness to others—as well as helping the person immunized.



BRING IT HOME

The Benefits of Voyager I Endure

While we applaud the technology spillovers of NASA's space projects, we should also acknowledge that those benefits are not shared equally. Economists like Tyler Cowen, a professor at George Mason University, are seeing increasing evidence of a widening gap between those who have access to rapidly improving technology, and those who do not. According to Cowen, author of the 2013 book, *Average Is Over: Powering America Beyond the Age of the Great Stagnation*, this inequality in access to technology and information is going to deepen the inequality in

skills, and ultimately, in wages and global standards of living.

Key Terms

external benefits (or positive externalities) beneficial spillovers to a third party or parties, who did not purchase the good or service that provided the externalities

free rider those who want others to pay for the public good and then plan to use the good themselves; if many people act as free riders, the public good may never be provided

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

nonexcludable when it is costly or impossible to exclude someone from using the good, and thus hard to charge for it

nonrivalrous even when one person uses the good, others can also use it

positive externalities beneficial spillovers to a third party or parties

private benefits the benefits a person who consumes a good or service receives, or a new product's benefits or process that a company invents that the company captures

private rates of return when the estimated rates of return go primarily to an individual; for example, earning interest on a savings account

public good good that is nonexcludable and non-rival, and thus is difficult for market producers to sell to individual consumers

social benefits the sum of private benefits and external benefits

social rate of return when the estimated rates of return go primarily to society; for example, providing free education

Key Concepts and Summary

13.1 Investments in Innovation

Competition creates pressure to innovate. However, if one can easily copy new inventions, then the original inventor loses the incentive to invest further in research and development. New technology often has positive externalities; that is, there are often spillovers from the invention of new technology that benefit firms other than the innovator. The social benefit of an invention, once the firm accounts for these spillovers, typically exceeds the private benefit to the inventor. If inventors could receive a greater share of the broader social benefits for their work, they would have a greater incentive to seek out new inventions.

13.2 How Governments Can Encourage Innovation

Public policy with regard to technology must often strike a balance. For example, patents provide an incentive for inventors, but they should be limited to genuinely new inventions and not extend forever.

Government has a variety of policy tools for increasing the rate of return for new technology and encouraging its development, including: direct government funding of R&D, tax incentives for R&D, protection of intellectual property, and forming cooperative relationships between universities and the private sector.

13.3 Public Goods

A public good has two key characteristics: it is nonexcludable and non-rival. Nonexcludable means that it is costly or impossible for one user to exclude others from using the good. Non-rival means that when one person uses the good, it does not prevent others from using it. Markets often have a difficult time producing public goods because free riders will attempt to use the public good without paying for it. One can overcome the free rider problem through measures to assure that users of the public good pay for it. Such measures include government actions, social pressures, and specific situations where markets have discovered a way to collect payments.

Self-Check Questions

1. Do market demand curves reflect positive externalities? Why or why not?

2. Suppose that Sony's R&D investment in digital devices has increased profits by 20%. Is this a private or social benefit?
3. The Gizmo Company is planning to develop new household gadgets. [Table 13.4](#) shows the company's demand for financial capital for research and development of these gadgets, based on expected rates of return from sales. Now, say that every investment would have an additional 5% social benefit—that is, an investment that pays at least a 6% return to the Gizmo Company will pay at least an 11% return for society as a whole; an investment that pays at least 7% for the Gizmo Company will pay at least 12% for society as a whole, and so on. Answer the questions that follow based on this information.

Estimated Rate of Return	Private profits of the firm from an R&D project (in \$ millions)
10%	\$100
9%	\$102
8%	\$108
7%	\$118
6%	\$133
5%	\$153
4%	\$183
3%	\$223

TABLE 13.4

- a. If the going interest rate is 9%, how much will Gizmo invest in R&D if it receives only the private benefits of this investment?
- b. Assume that the interest rate is still 9%. How much will the firm invest if it also receives the social benefits of its investment? (Add an additional 5% return on all levels of investment.)
4. The Junkbuyers Company travels from home to home, looking for opportunities to buy items that would otherwise end up with the garbage, but which the company can resell or recycle. Which will be larger, the private or the social benefits?
5. When residents in a neighborhood tidy it and keep it neat, there are a number of positive spillovers: higher property values, less crime, happier residents. What types of government policies can encourage neighborhoods to clean up?
6. Education provides both private benefits to those who receive it and broader social benefits for the economy as a whole. Think about the types of policies a government can follow to address the issue of positive spillovers in technology and then suggest a parallel set of policies that governments could follow for addressing positive externalities in education.
7. Which of the following goods or services are nonexcludable?
 - a. police protection
 - b. streaming music from satellite transmission programs
 - c. roads
 - d. primary education
 - e. cell phone service

8. Are the following goods non-rival in consumption?
- slice of pizza
 - laptop computer
 - public radio
 - ice cream cone

Review Questions

9. In what ways do company investments in research and development create positive externalities?
10. Will the demand for borrowing and investing in R&D be higher or lower if there are no external benefits?
11. Why might private markets tend to provide too few incentives for the development of new technology?
12. What can government do to encourage the development of new technology?
13. What are the two key characteristics of public goods?
14. Name two public goods and explain why they are public goods.
15. What is the free rider problem?
16. Explain why the federal government funds national defense.

Critical Thinking Questions

17. Can a company be guaranteed all of the social benefits of a new invention? Why or why not?
18. Is it inevitable that government must become involved in supporting investments in new technology?
19. How do public television stations, like PBS, try to overcome the free rider problem?
20. Why is a football game on ESPN a quasi-public good but a game on the NBC, CBS, or ABC is a public good?
21. Provide two examples of goods/services that are classified as private goods/services even though they are provided by a federal government.
22. Radio stations, tornado sirens, light houses, and street lights are all public goods in that all are nonrivalrous and nonexclusionary. Therefore why does the government provide tornado sirens, street lights and light houses but not radio stations (other than PBS stations)?

Problems

- 23.** HighFlyer Airlines wants to build new airplanes with greatly increased cabin space. This will allow HighFlyer Airlines to give passengers more comfort and sell more tickets at a higher price. However, redesigning the cabin means rethinking many other elements of the airplane as well, like engine and luggage placement and the most efficient shape of the plane for moving through the air. HighFlyer Airlines has developed a list of possible methods to increase cabin space, along with estimates of how these approaches would affect the plane's operating costs and ticket sales. Based on these estimates, [Table 13.5](#) shows the value of R&D projects that provide at least a certain private rate of return. Column 1 = Private Rate of Return. Column 2 = Value of R&D Projects that Return at Least the Private Rate of Return to HighFlyer Airlines. Use the data to answer the following questions.

Private Rate of Return	Value of R&D
12%	\$100
10%	\$200
8%	\$300
6%	\$400
4%	\$500

TABLE 13.5

- If the opportunity cost of financial capital for HighFlyer Airlines is 6%, how much should the firm invest in R&D?
 - Assume that the social rate of return for R&D is an additional 2% on top of the private return; that is, an R&D investment that had a 7% private return to HighFlyer Airlines would have a 9% social return. How much investment is socially optimal at the 6% interest rate?
- 24.** Assume that the marginal private costs of a firm producing fuel-efficient cars is greater than the marginal social costs. Assume that the marginal private benefits of a firm producing fuel-efficient cars is the same as the marginal social benefits. Discuss one way that the government can try to increase production and sales of fuel efficient cars to the socially desirable amount. *Hint:* the government is trying to affect production through costs, not benefits.
- 25.** Becky and Sarah are sisters who share a room. Their room can easily get messy, and their parents are always telling them to tidy it. Here are the costs and benefits to both Becky and Sarah, of taking the time to clean their room: If both Becky and Sarah clean, they each spends two hours and get a clean room. If Becky decides not to clean and Sarah does all the cleaning, then Sarah spends 10 hours cleaning (Becky spends 0) but Sarah is exhausted. The same would occur for Becky if Sarah decided not to clean—Becky spends 10 hours and becomes exhausted. If both girls decide not to clean, they both have a dirty room.
- What is the best outcome for Becky and Sarah? What is the worst outcome? (It would help you to construct a prisoner's dilemma table.)
 - Unfortunately, we know that the optimal outcome will most likely not happen, and that the sisters probably will choose the worst one instead. Explain what it is about Becky's and Sarah's reasoning that will lead them both to choose the worst outcome.



FIGURE 14.1 What determines incomes? In the U.S., income is primarily based on one's value to an employer, which depends in part on education. (Credit: modification of work by AFL-CIO America's Unions/Flickr Creative Commons and COD Newsroom/Flickr Creative Commons)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- The theory of labor markets
- How wages are determined in an imperfectly competitive labor market
- How unions affect wages and employment
- How labor market outcomes are determined under Bilateral Monopoly
- Theories of Employment Discrimination, and
- How Immigration affects labor market outcomes

Introduction to Labor Markets and Income



BRING IT HOME

The Increasing Value of a College Degree

Working your way through college used to be fairly common in the United States. According to a 2015 study by the Georgetown Center on Education and the Workforce, 40% of college students work 30 hours or more per week.

At the same time, the cost of college seems to rise every year. The data show that between the 2000–2001 academic year and the 2019–2020 academic year, the cost of tuition, fees, and room and board has slightly more than doubled for private four-year colleges, and has increased by a factor of almost 2.5 for public four-year colleges. Thus, even full time employment may not be enough to cover college expenses anymore. Working full time at minimum wage—40 hours per week, 52 weeks per year—earns \$15,080 before taxes, which is substantially less

than the more than \$25,000 estimated as the average cost in 2022 for a year of college at a public university. The result of these costs is that student loan debt topped \$1.3 trillion this year.

Despite these disheartening figures, the value of a bachelor's degree has never been higher. How do we explain this? This chapter will tell us.

In a market economy like the United States, income comes from ownership of the means of production: resources or assets. More precisely, one's income is a function of two things: the quantity of each resource one owns, and the value society places on those resources. Recall from the chapter on [Production, Costs, and Industry Structure](#), each factor of production has an associated factor payment. For the majority of us, the most important resource we own is our labor. Thus, most of our income is wages, salaries, commissions, tips and other types of labor income. Your labor income depends on how many hours you work and the wage rate an employer will pay you for those hours. At the same time, some people own real estate, which they can either use themselves or rent out to other users. Some people have financial assets like bank accounts, stocks and bonds, for which they earn interest, dividends or some other form of income.

Each of these factor payments, like wages for labor and interest for financial capital, is determined in their respective factor markets. For the rest of this chapter, we will focus on labor markets, but other factor markets operate similarly. Later in Chapter 17 we will describe how this works for financial capital.

14.1 The Theory of Labor Markets

LEARNING OBJECTIVES

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

- Describe the demand for labor in perfectly competitive output markets
- Describe the demand for labor in imperfectly competitive output markets
- Identify what determines the going market rate for labor



CLEAR IT UP

What is the labor market?

The labor market is the term that economists use for all the different markets for labor. There is no single labor market. Rather, there is a different market for every different type of labor. Labor differs by type of work (e.g. retail sales vs. scientist), skill level (entry level or more experienced), and location (the market for administrative assistants is probably more local or regional than the market for university presidents). While each labor market is different, they all tend to operate in similar ways. For example, when wages go up in one labor market, they tend to go up in others too. When economists talk about the labor market, they are describing these similarities.

The labor market, like all markets, has a demand and a supply. Why do firms demand labor? Why is an employer willing to pay you for your labor? It's not because the employer likes you or is socially conscious. Rather, it's because your labor is worth something to the employer--your work brings in revenues to the firm. How much is an employer willing to pay? That depends on the skills and experience you bring to the firm.

If a firm wants to maximize profits, it will never pay more (in terms of wages and benefits) for a worker than the value of their marginal productivity to the firm. We call this the **first rule of labor markets**.

Suppose a worker can produce two widgets per hour and the firm can sell each widget for \$4 each. Then the worker is generating \$8 per hour in revenues to the firm, and a profit-maximizing employer will pay the worker up to, but no more than, \$8 per hour, because that is what the worker is worth to the firm.

Recall the definition of marginal product. Marginal product is the additional output a firm can produce by adding one more worker to the production process. Since employers often hire labor by the hour, we'll define

marginal product as the additional output the firm produces by adding one more worker hour to the production process. In this chapter, we assume that workers in a particular labor market are homogeneous—they have the same background, experience and skills and they put in the same amount of effort. Thus, marginal product depends on the capital and technology with which workers have to work.

A typist can type more pages per hour with an electric typewriter than a manual typewriter, and the typist can type even more pages per hour with a personal computer and word processing software. A ditch digger can dig more cubic feet of dirt in an hour with a backhoe than with a shovel.

Thus, we can define the demand for labor as the marginal product of labor times the value of that output to the firm.

# Workers (L)	1	2	3	4
MP_L	4	3	2	1

TABLE 14.1 Marginal Product of Labor

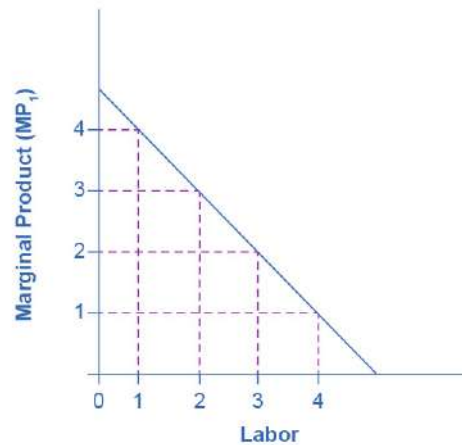


FIGURE 14.2 Marginal Product of Labor Because of fixed capital, the marginal product of labor declines as the employer hires additional workers.

On what does the value of each worker's marginal product depend? If we assume that the employer sells its output in a perfectly competitive market, the value of each worker's output will be the market price of the product. Thus,

$$\text{Demand for Labor} = MP_L \times P = \text{Value of the Marginal Product of Labor}$$

We show this in [Table 14.2](#), which is an expanded version of [Table 14.1](#)

# Workers (L)	1	2	3	4
MP_L	4	3	2	1
Price of Output	\$4	\$4	\$4	\$4
VMP_L	\$16	\$12	\$8	\$4

TABLE 14.2 Value of the Marginal Product of Labor

Note that the value of each additional worker is less than the value of the ones who came before.

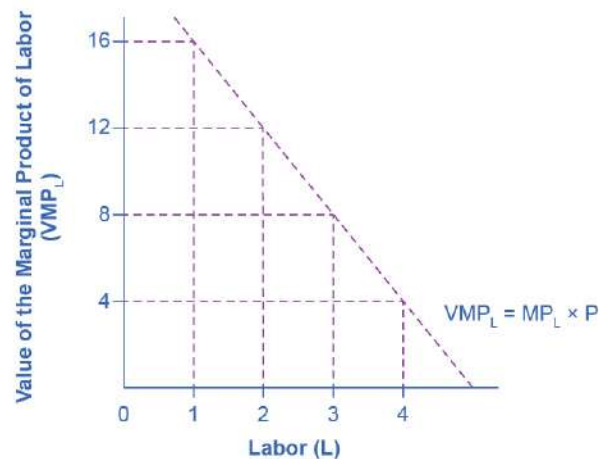


FIGURE 14.3 Value of the Marginal Product of Labor For firms operating in a competitive output market, the value of additional output sold is the price the firms receive for the output. Since MP_L declines with additional labor employed, while that marginal product is worth the market price, the value of the marginal product declines as employment increases.

Demand for Labor in Perfectly Competitive Output Markets

The question for any firm is how much labor to hire.

We can define a **Perfectly Competitive Labor Market** as one where firms can hire all the labor they want at the going market wage. Think about secretaries in a large city. Employers who need secretaries can probably hire as many as they need if they pay the going wage rate.

Graphically, this means that firms face a horizontal supply curve for labor, as Figure 14.3 shows.

Given the market wage, profit maximizing firms hire workers up to the point where: $W_{\text{mkt}} = VMP_L$

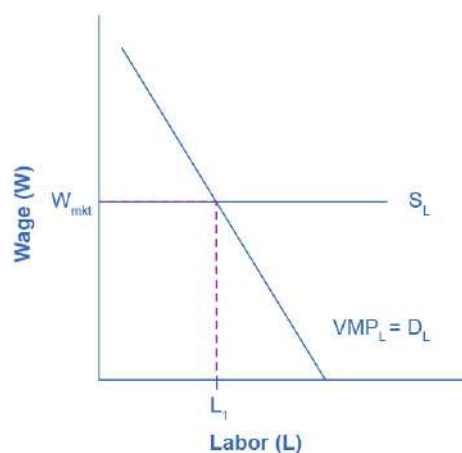


FIGURE 14.4 Equilibrium Employment for Firms in a Competitive Labor Market In a perfectly competitive labor market, firms can hire all the labor they want at the going market wage. Therefore, they hire workers up to the point L_1 where the going market wage equals the value of the marginal product of labor.



CLEAR IT UP

Derived Demand

Economists describe the demand for inputs like labor as a **derived demand**. Since the demand for labor is $MPL \times P$, it is dependent on the demand for the product the firm is producing. We show this by the P term in the demand for labor. An increase in demand for the firm's product drives up the product's price, which increases the firm's demand for labor. Thus, we derive the demand for labor from the demand for the firm's output.

Demand for Labor in Imperfectly Competitive Output Markets

If the employer does not sell its output in a perfectly competitive industry, they face a downward sloping demand curve for output, which means that in order to sell additional output the firm must lower its price. This is true if the firm is a monopoly, but it's also true if the firm is an oligopoly or monopolistically competitive. In this situation, the value of a worker's marginal product is the marginal revenue, not the price. Thus, the demand for labor is the marginal product times the marginal revenue.

The Demand for Labor = $MP_L \times MR$ = Marginal Revenue Product

# Workers (L)	1	2	3	4
MP_L	4	3	2	1
Marginal Revenue	\$4	\$3	\$2	\$1
MRP_L	\$16	\$9	\$4	\$1

TABLE 14.3 Marginal Revenue Product

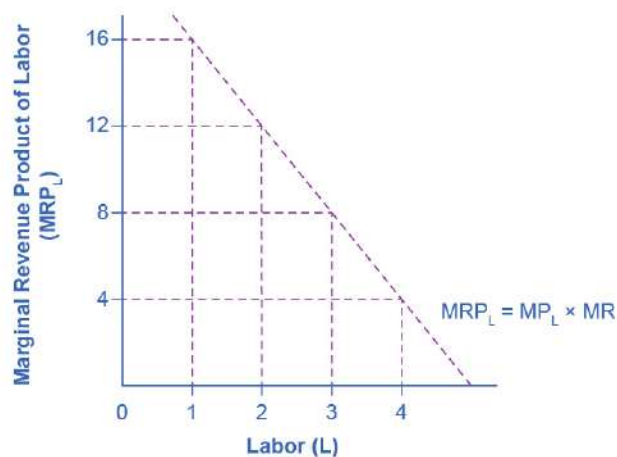


FIGURE 14.5 Marginal Revenue Product For firms with some market power in their output market, the value of additional output sold is the firm's marginal revenue. Since MP_L declines with additional labor employed and since MR declines with additional output sold, the firm's marginal revenue declines as employment increases.

Everything else remains the same as we described above in the discussion of the labor demand in perfectly competitive labor markets. Given the market wage, profit-maximizing firms will hire workers up to the point where the market wage equals the marginal revenue product, as [Figure 14.6](#) shows.

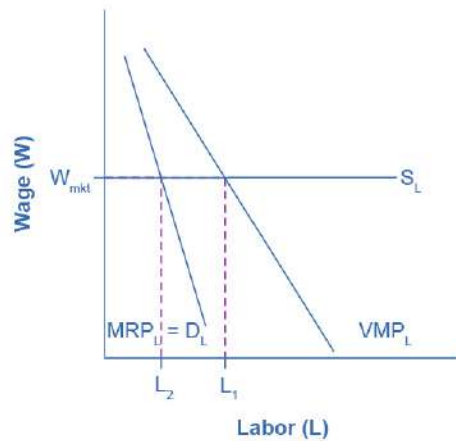


FIGURE 14.6 Equilibrium Level of Employment for Firms with Market Power For firms with market power in their output market, they choose the number of workers, L_2 , where the going market wage equals the firm's marginal revenue product. Note that since marginal revenue is less than price, the demand for labor for a firm which has market power in its output market is less than the demand for labor (L_1) for a perfectly competitive firm. As a result, employment will be lower in an imperfectly competitive industry than in a perfectly competitive industry.



CLEAR IT UP

Do Profit Maximizing Employers Exploit Labor?

If you look back at [Figure 14.4](#), you will see that the firm pays only the last worker it hires what they're worth to the firm. Every other worker brings in more revenue than the firm pays them. This has sometimes led to the claim that employers exploit workers because they do not pay workers what they are worth. Let's think about this claim. The first worker is worth \$ x to the firm, and the second worker is worth \$ y , but why are they worth that much? It is because of the capital and technology with which they work. The difference between workers' worth and their compensation goes to pay for the capital and technology, without which the workers wouldn't have a job. The difference also goes to the employer's profit, without which the firm would close and workers wouldn't have a job. The firm may be earning excessive profits, but that is a different topic of discussion.

What Determines the Going Market Wage Rate?

In the chapter on [Labor and Financial Markets](#), we learned that the labor market has demand and supply curves like other markets. The demand for labor curve is a downward sloping function of the wage rate. The market demand for labor is the horizontal sum of all firms' demands for labor. The supply of labor curve is an upward sloping function of the wage rate. This is because if wages for a particular type of labor increase in a particular labor market, people with appropriate skills may change jobs, and vacancies will attract people from outside the geographic area. The market supply of labor is the horizontal summation of all individuals' supplies of labor.

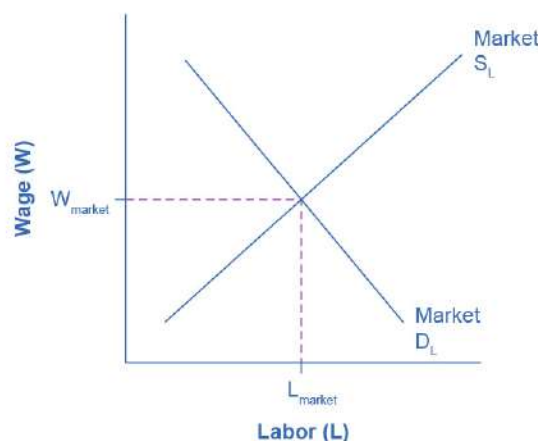


FIGURE 14.7 The Market Wage Rate In a competitive labor market, the equilibrium wage and employment level are determined where the market demand for labor equals the market supply of labor.

Like all equilibrium prices, the market wage rate is determined through the interaction of supply and demand in the labor market. Thus, we can see in [Figure 14.7](#) for competitive markets the wage rate and number of workers hired.

The FRED database has a great deal of data on labor markets, starting at [the wage rate and number of workers hired \(https://openstax.org/l/cat10\)](https://openstax.org/l/cat10).

The United States Census Bureau for the Bureau of Labor Statistics publishes *The Current Population Survey*, which is a monthly survey of households (you can find a link to it by going to the FRED database found in the previous link), which provides data on labor supply, including numerous measures of the labor force size (disaggregated by age, gender and educational attainment), labor force participation rates for different demographic groups, and employment. It also includes more than 3,500 measures of earnings by different demographic groups.

The Current Employment Statistics, which is a survey of businesses, offers alternative estimates of employment across all sectors of the economy.

The FRED database, found in the previous link, also has a link labeled "Productivity and Costs" has a wide range of data on productivity, labor costs, and profits across the business sector.

14.2 Wages and Employment in an Imperfectly Competitive Labor Market

LEARNING OBJECTIVES

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

- Define monopsony power
- Explain how imperfectly competitive labor markets determine wages and employment, where employers have market power

In the chapters on market structure, we observed that while economists use the theory of perfect competition as an ideal case of market structure, there are very few examples of perfectly competitive industries in the real world. What about labor markets? How many labor markets are perfectly competitive? There are probably more examples of perfectly competitive labor markets than perfectly competitive product markets, but that doesn't mean that all labor markets are competitive.

When a job applicant is bargaining with an employer for a position, the applicant is often at a disadvantage—needing the job more than the employer needs that particular applicant. John Bates Clark (1847–1938), often named as the first great American economist, wrote in 1907: “In the making of the wages contract the individual laborer is always at a disadvantage. He has something which he is obliged to sell and

which his employer is not obliged to take, since he [that is, the employer] can reject single men with impunity.”

To give workers more power, the U.S. government has passed, in response to years of labor protests, a number of laws to create a more equal balance of power between workers and employers. These laws include some of the following:

- Setting minimum hourly wages
- Setting maximum hours of work (at least before employers pay overtime rates)
- Prohibiting child labor
- Regulating health and safety conditions in the workplace
- Preventing discrimination on the basis of race, ethnicity, gender, sexual orientation, and age
- Requiring employers to provide family leave
- Requiring employers to give advance notice of layoffs
- Covering workers with unemployment insurance
- Setting a limit on the number of immigrant workers from other countries

[Table 14.4](#) lists some prominent U.S. workplace protection laws. Many of the laws listed in the table were only the start of labor market regulations in these areas and have been followed, over time, by other related laws, regulations, and court rulings.

Law	Protection
National Labor-Management Relations Act of 1935 (the “Wagner Act”)	Establishes procedures for establishing a union that firms are obligated to follow; sets up the National Labor Relations Board for deciding disputes
Social Security Act of 1935	Under Title III, establishes a state-run system of unemployment insurance, in which workers pay into a state fund when they are employed and received benefits for a time when they are unemployed
Fair Labor Standards Act of 1938	Establishes the minimum wage, limits on child labor, and rules requiring payment of overtime pay for those in jobs that are paid by the hour and exceed 40 hours per week
Taft-Hartley Act of 1947	Allows states to decide whether all workers at a firm can be required to join a union as a condition of employment; in the case of a disruptive union strike, permits the president to declare a “cooling-off period” during which workers have to return to work
Civil Rights Act of 1964	Title VII of the Act prohibits discrimination in employment on the basis of race, gender, national origin, religion, or sexual orientation
Occupational Health and Safety Act of 1970	Creates the Occupational Safety and Health Administration (OSHA), which protects workers from physical harm in the workplace
Employee Retirement and Income Security Act of 1974	Regulates employee pension rules and benefits
Pregnancy Discrimination Act of 1978	Prohibits discrimination against women in the workplace who are planning to get pregnant or who are returning to work after pregnancy

TABLE 14.4 Prominent U.S. Workplace Protection Laws

Law	Protection
Immigration Reform and Control Act of 1986	Prohibits hiring of illegal immigrants; requires employers to ask for proof of citizenship; protects rights of legal immigrants
Worker Adjustment and Retraining Notification Act of 1988	Requires employers with more than 100 employees to provide written notice 60 days before plant closings or large layoffs
Americans with Disabilities Act of 1990	Prohibits discrimination against those with disabilities and requires reasonable accommodations for them on the job
Family and Medical Leave Act of 1993	Allows employees to take up to 12 weeks of unpaid leave per year for family reasons, including birth or family illness
Pension Protection Act of 2006	Penalizes firms for underfunding their pension plans and gives employees more information about their pension accounts
Lilly Ledbetter Fair Pay Act of 2009	Restores protection for pay discrimination claims on the basis of sex, race, national origin, age, religion, or disability

TABLE 14.4 Prominent U.S. Workplace Protection Laws

There are two sources of imperfect competition in labor markets. These are demand side sources, that is, labor market power by employers, and supply side sources: labor market power by employees. In this section we will discuss the former. In the next section we will discuss the latter.

A competitive labor market is one where there are many potential employers for a given type of worker, say a secretary or an accountant. Suppose there is only one employer in a labor market. Because that employer has no direct competition in hiring, if they offer lower wages than would exist in a competitive market, employees will have few options. If they want a job, they must accept the offered wage rate. Since the employer is exploiting its market power, we call the firm a **monopsony**, a term introduced and widely discussed by Joan Robinson (though she credited scholar Bertrand Hallward with invention of the word). The classical example of monopsony is the sole coal company in a West Virginia town. If coal miners want to work, they must accept what the coal company is paying. This is not the only example of monopsony. Think about surgical nurses in a town with only one hospital. A situation in which employers have at least some market power over potential employees is not that unusual. After all, most firms have many employees while there is only one employer. Thus, even if there is some competition for workers, it may not feel that way to potential employees unless they do their research and find the opposite.

How does market power by an employer affect labor market outcomes? Intuitively, one might think that wages will be lower than in a competitive labor market. Let's prove it. We will tell the story for a monopsonist, but the results will be qualitatively similar, although less extreme, for any firm with labor market power.

Think back to monopoly. The good news for the firm is that because the monopolist is the sole supplier in the market, it can charge any price it wishes. The bad news is that if it wants to sell a greater quantity of output, it must lower the price it charges. Monopsony is analogous. Because the monopsonist is the sole employer in a labor market, it can offer any wage that it wishes. However, because they face the market supply curve for labor, if they want to hire more workers, they must raise the wage they pay. This creates a quandary, which we can understand by introducing a new concept: the marginal cost of labor. The **marginal cost of labor** is the cost to the firm of hiring one more worker. However, here is the thing: we assume that the firm is determining how many workers to hire in total. They are not hiring sequentially. Let's look how this plays out with the

example in [Table 14.5](#).

Supply of Labor	1	2	3	4	5
Wage Rate	\$1 per hour	\$2 per hour	\$3 per hour	\$4 per hour	\$5 per hour
Total Cost of Labor	\$1	\$4	\$9	\$16	\$25
Marginal Cost of Labor	\$1	\$3	\$5	\$7	\$9

TABLE 14.5 The Marginal Cost of Labor

There are a couple of things to notice from the table. First, the marginal cost increases faster than the wage rate. In fact, for any number of workers more than one, the marginal cost of labor is greater than the wage. This is because to hire one more worker requires paying a higher wage rate, not just for the new worker but for all the previous hires also. We can see this graphically in Figure 14.7.

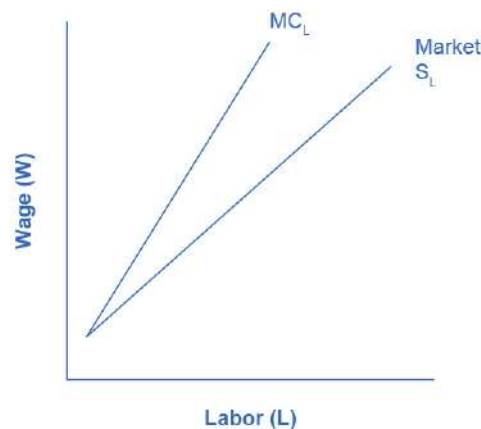


FIGURE 14.8 The Marginal Cost of Labor Since monopsonies are the sole demander for labor, they face the market supply curve for labor. In order to increase employment they must raise the wage they pay not just for new workers, but for all the existing workers they could have hired at the previous lower wage. As a result, the marginal cost of hiring additional labor is greater than the wage, and thus for any level of employment (above the first worker), MC_L is above the Market Supply of Labor.

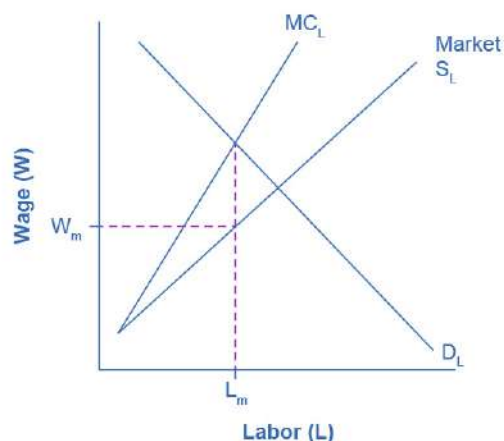


FIGURE 14.9 Labor Market Outcomes Under Monopsony A monopsony will hire workers up to the point L_m where its demand for labor equals the marginal cost of additional labor, paying the wage W_m given by the supply curve of

labor necessary to obtain L_m workers.

If the firm wants to maximize profits, it will hire labor up to the point L_m where $D_L = VMP$ (or MRP) = MC_L , as [Figure 14.9](#) shows. Then, the supply curve for labor shows the wage the firm will have to pay to attract L_m workers. Graphically, we can draw a vertical line up from L_m to the Supply Curve for the label and then read the wage W_m off the vertical axis to the left.

How does this outcome compare to what would occur in a perfectly competitive market? A competitive market would operate where $D_L = S_L$, hiring L_c workers and paying W_c wage. In other words, under monopsony employers hire fewer workers and pay a lower wage. While pure monopsony may be rare, many employers have some degree of market power in labor markets. The outcomes for those employers will be qualitatively similar though not as extreme as monopsony.

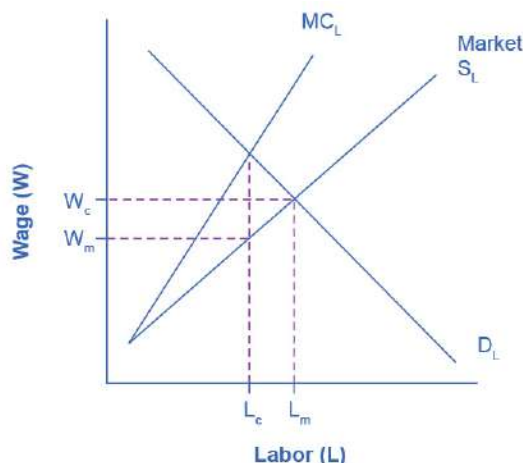


FIGURE 14.10 Comparison of labor market outcomes: Monopsony vs. Perfect Competition A monopsony hires fewer workers (L_m) than would be hired in a competitive labor market (L_c). In exploiting its market power, the monopsony can also pay a lower wage (W_m) than workers would earn in a competitive labor market (W_c).

14.3 Market Power on the Supply Side of Labor Markets: Unions

LEARNING OBJECTIVES

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

- Explain the concept of labor unions, including membership levels and wages
- Evaluate arguments for and against labor unions
- Analyze reasons for the decline in U.S. union membership

A labor union is an organization of workers that negotiates with employers over wages and working conditions. A labor union seeks to change the balance of power between employers and workers by requiring employers to deal with workers collectively, rather than as individuals. As such, a labor union operates like a monopoly in a labor market. We sometimes call negotiations between unions and firms **collective bargaining**.

The subject of labor unions can be controversial. Supporters of labor unions view them as the workers' primary line of defense against efforts by profit-seeking firms to hold down wages and benefits. Critics of labor unions view them as having a tendency to grab as much as they can in the short term, even if it means injuring workers in the long run by driving firms into bankruptcy or by blocking the new technologies and production methods that lead to economic growth. We will start with some facts about union membership in the United States.

Facts about Union Membership and Pay

According to the U.S. Bureau of Labor and Statistics, about 10.3% of all U.S. workers belong to unions. This represents nearly a 50% reduction since 1983 (the earliest year for which comparable data are available),

when union members were 20.1% of all workers. Following are some facts about unions for 2021 (note that we are using the population categories and group names utilized in the data collection and publication):

- 10.6% of U.S. male workers belong to unions; 9.9% of female workers do
- 10.7% of White workers, 12.3% of Black workers, and 9.8 % of Hispanic workers belong to unions
- 11.8% of full-time workers and 5.7% of part-time workers are union members
- 4.4% of workers ages 16–24 belong to unions, as do 13.2% of workers ages 45–54
- Occupations in which relatively high percentages of workers belong to unions are the federal government (26.0% belong to a union), state government (29.9%), local government (41.7%); transportation and utilities (17.6%); natural resources, construction, and maintenance (15.9%); and production, transportation, and material moving (13.3%)
- Occupations that have relatively low percentages of unionized workers are agricultural workers (1.7%), financial services (1.9%), professional and business services (2.2%), leisure and hospitality (2.2%), and wholesale and retail trade (4.5%)

In summary, the percentage of workers belonging to a union is higher for men than women; higher for Black than for White or Hispanic people; higher for the 45–64 age range; and higher among workers in government and manufacturing than workers in agriculture or service-oriented jobs. [Table 14.6](#) lists the largest U.S. labor unions and their membership.

Union	Membership
National Education Association (NEA)	3.0 million
Service Employees International Union (SEIU)	2.0 million
American Federation of Teachers (AFT)	1.7 million
International Brotherhood of Teamsters (IBT)	1.4 million
The American Federation of State, County, and Municipal Workers (AFSCME)	1.6 million
United Food and Commercial Workers International Union	1.3 million
International Brotherhood of Electrical Workers (IBEW)	775,000
United Steelworkers	625,000
International Association of Machinists and Aerospace Workers	569,000
International Union, United Automobile, Aerospace and Agricultural Implement Workers of America (UAW)	408,000

TABLE 14.6 The Largest American Unions in 2021 (Source: U.S. Department of Labor and individual union websites)

In terms of pay, benefits, and hiring, U.S. unions offer a good news/bad news story. The good news for unions and their members is that their members earn about 20% more than nonunion workers, even after adjusting for factors such as years of work experience and education level. The bad news for unions is that the share of U.S. workers who belong to a labor union has been steadily declining for 50 years, as [Figure 14.11](#) shows. About one-quarter of all U.S. workers belonged to a union in the mid-1950s, but only 10.3% of U.S. workers are union members today. If you leave out government workers (which includes teachers in public schools), only 6.1% of the workers employed by private firms now work for a union.

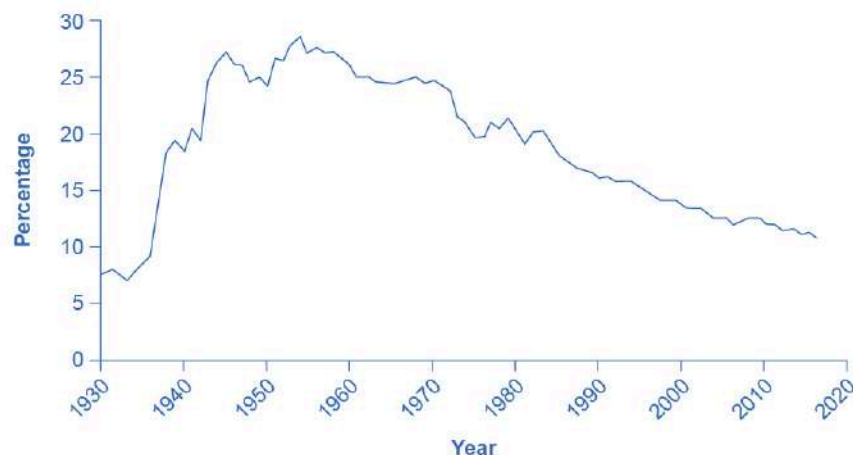


FIGURE 14.11 Percentage of Wage and Salary Workers Who Are Union Members The share of wage and salary workers who belong to unions rose sharply in the 1930s and 1940s, but has tailed off since then to 10.3% of all workers in 2021.

The following section analyzes the higher pay union workers receive compared the pay rates for nonunion workers. The section after that analyzes declining union membership levels. An overview of these two issues will allow us to discuss many aspects of how unions work.

Higher Wages for Union Workers

How does a union affect wages and employment? Because a union is the sole supplier of labor, it can act like a monopoly and ask for whatever wage rate it can obtain for its workers. If employers need workers, they have to meet the union's wage demand.

What are the limits on how much higher pay union workers can receive? To analyze these questions, let's consider a situation where all firms in an industry must negotiate with a single union, and no firm is allowed to hire nonunion labor. If no labor union existed in this market, then equilibrium (E) in the labor market would occur at the intersection of the demand for labor (D) and the supply of labor (S) as we see in [Figure 14.12](#). This is the same result as we showed in Figure 14.6 above. The union can, however, threaten that, unless firms agree to the wages they demand, the workers will strike. As a result, the labor union manages to achieve, through negotiations with the firms, a union wage of W_u for its members, above what the equilibrium wage would otherwise have been.

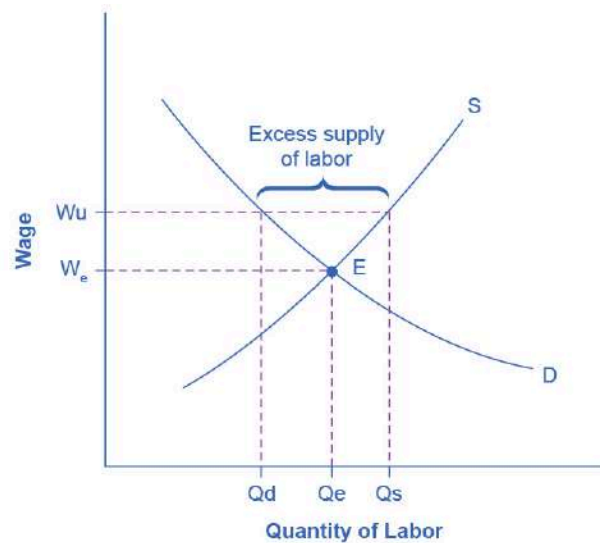


FIGURE 14.12 Union Wage Negotiations Without a union, the equilibrium at E would have involved the wage W_e and the quantity of labor Q_e . However, the union is able to use its bargaining power to raise the wage to W_u . The result is an excess supply of labor for union jobs. That is, a quantity of labor supplied, Q_s is greater than firms' quantity demanded for labor, Q_d .

This labor market situation resembles what a monopoly firm does in selling a product, but in this case a union is a monopoly selling labor to firms. At the higher union wage W_u , the firms in this industry will hire less labor than they would have hired in equilibrium. Moreover, an excess supply of workers want union jobs, but firms will not be hiring for such jobs.

From the union point of view, workers who receive higher wages are better off. However, notice that the quantity of workers (Q_d) hired at the union wage W_u is smaller than the quantity Q_e that the firm would have hired at the original equilibrium wage. A sensible union must recognize that when it pushes up the wage, it also reduces the firms' incentive to hire. This situation does not necessarily mean that union workers are fired. Instead, it may be that when union workers move on to other jobs or retire, they are not always replaced, or perhaps when a firm expands production, it expands employment somewhat less with a higher union wage than it would have done with the lower equilibrium wage. Other situations could be that a firm decides to purchase inputs from nonunion producers, rather than producing them with its own highly paid unionized workers, or perhaps the firm moves or opens a new facility in a state or country where unions are less powerful.

From the firm's point of view, the key question is whether union workers' higher wages are matched by higher productivity. If so, then the firm can afford to pay the higher union wages and, the demand curve for "unionized" labor could actually shift to the right. This could reduce the job losses as the equilibrium employment level shifts to the right and the difference between the equilibrium and the union wages will have been reduced. If worker unionization does not increase productivity, then the higher union wage will cause lower profits or losses for the firm.

Union workers might have higher productivity than nonunion workers for a number of reasons. First, higher wages may elicit higher productivity. Second, union workers tend to stay longer at a given job, a trend that reduces the employer's costs for training and hiring and results in workers with more years of experience. Many unions also offer job training and apprenticeship programs.

In addition, firms that are confronted with union demands for higher wages may choose production methods that involve more physical capital and less labor, resulting in increased labor productivity. [Table 14.7](#) provides an example. Assume that a firm can produce a home exercise cycle with three different combinations of labor and manufacturing equipment. Say that the firm pays labor \$16 an hour (including benefits) and the machines

for manufacturing cost \$200 each. Under these circumstances, the total cost of producing a home exercise cycle will be lowest if the firm adopts the plan of 50 hours of labor and one machine, as the table shows. Now, suppose that a union negotiates a wage of \$20 an hour including benefits. In this case, it makes no difference to the firm whether it uses more hours of labor and fewer machines or less labor and more machines, although it might prefer to use more machines and to hire fewer union workers. (After all, machines never threaten to strike—but they do not buy the final product or service either.)

In the final column of the table, the wage has risen to \$24 an hour. In this case, the firm clearly has an incentive for using the plan that involves paying for fewer hours of labor and using three machines. If management responds to union demands for higher wages by investing more in machinery, then union workers can be more productive because they are working with more or better physical capital equipment than the typical nonunion worker. However, the firm will need to hire fewer workers.

Hours of Labor	Number of Machines	Cost of Labor + Cost of Machine \$16/hour	Cost of Labor + Cost of Machine \$20/hour	Cost of Labor + Cost of Machine \$24/hour
30	3	$\$480 + \$600 = \$1,080$	$\$600 + \$600 = \$1,200$	$\$720 + \$600 = \$1,320$
40	2	$\$640 + \$400 = \$1,040$	$\$800 + \$400 = \$1,200$	$\$960 + \$400 = \$1,360$
50	1	$\$800 + \$200 = \$1,000$	$\$1,000 + \$200 = \$1,200$	$\$1,200 + \$200 = \$1,400$

TABLE 14.7 Three Production Choices to Manufacture a Home Exercise Cycle

In some cases, unions have discouraged the use of labor-saving physical capital equipment—out of the reasonable fear that new machinery will reduce the number of union jobs. For example, in 2015, the union representing longshoremen who unload ships and the firms that operate shipping companies and port facilities staged a work stoppage that shut down the ports on the western coast of the United States. Two key issues in the dispute were the desire of the shipping companies and port operators to use handheld scanners for record-keeping and computer-operated cabs for loading and unloading ships—changes which the union opposed, along with overtime pay. President Obama threatened to use the Labor Management Relations Act of 1947—commonly known as the Taft-Hartley Act—where a court can impose an 80-day “cooling-off period” in order to allow time for negotiations to proceed without the threat of a work stoppage. Federal mediators were called in, and the two sides agreed to a deal in February 2015. The ultimate agreement allowed the new technologies, but also kept wages, health, and pension benefits high for workers. In the past, presidential use of the Taft-Hartley Act sometimes has made labor negotiations more bitter and argumentative but, in this case, it seems to have smoothed the road to an agreement.

In other instances, unions have proved quite willing to adopt new technologies. In one prominent example, during the 1950s and 1960s, the United Mineworkers union demanded that mining companies install labor-saving machinery in the mines. The mineworkers’ union realized that over time, the new machines would reduce the number of jobs in the mines, but the union leaders also knew that the mine owners would have to pay higher wages if the workers became more productive, and mechanization was a necessary step toward greater productivity.

In fact, in some cases union workers may be more willing to accept new technology than nonunion workers, because the union workers believe that the union will negotiate to protect their jobs and wages, whereas nonunion workers may be more concerned that the new technology will replace their jobs. In addition, union workers, who typically have higher job market experience and training, are likely to suffer less and benefit more than non-union workers from the introduction of new technology. Overall, it is hard to make a definitive case that union workers as a group are always either more or less welcoming to new technology than are nonunion workers.

The Decline in U.S. Union Membership

The proportion of U.S. workers belonging to unions has declined dramatically since the early 1950s. Economists have offered a number of possible explanations:

- The shift from manufacturing to service industries
- The force of globalization and increased competition from foreign producers
- A reduced desire for unions because of the workplace protection laws now in place
- U.S. legal environment that makes it relatively more difficult for unions to organize workers and expand their membership

Let's discuss each of these four explanations in more detail.

A first possible explanation for the decline in the share of U.S. workers belonging to unions involves the patterns of job growth in the manufacturing and service sectors of the economy as [Figure 14.13](#) shows. The U.S. economy had about 15 million manufacturing jobs in 1960. This total rose to 19 million by the late 1970s and then declined to 17 million in 2013. Meanwhile, the number of jobs in service industries (including government employment) rose from 35 million in 1960 to over 118 million by 2013, according to the Bureau of Labor Statistics. Because over time unions were stronger in manufacturing than in service industries, the growth in jobs was not happening where the unions were. It is interesting to note that government workers comprise several of the biggest unions in the country, including the **American Federation of State, County and Municipal Employees (AFSCME)**; the **Service Employees International Union**; and the **National Education Association**. [Table 14.8](#) lists the membership of each of these unions. Outside of government employees, however, unions have not had great success in organizing the service sector.

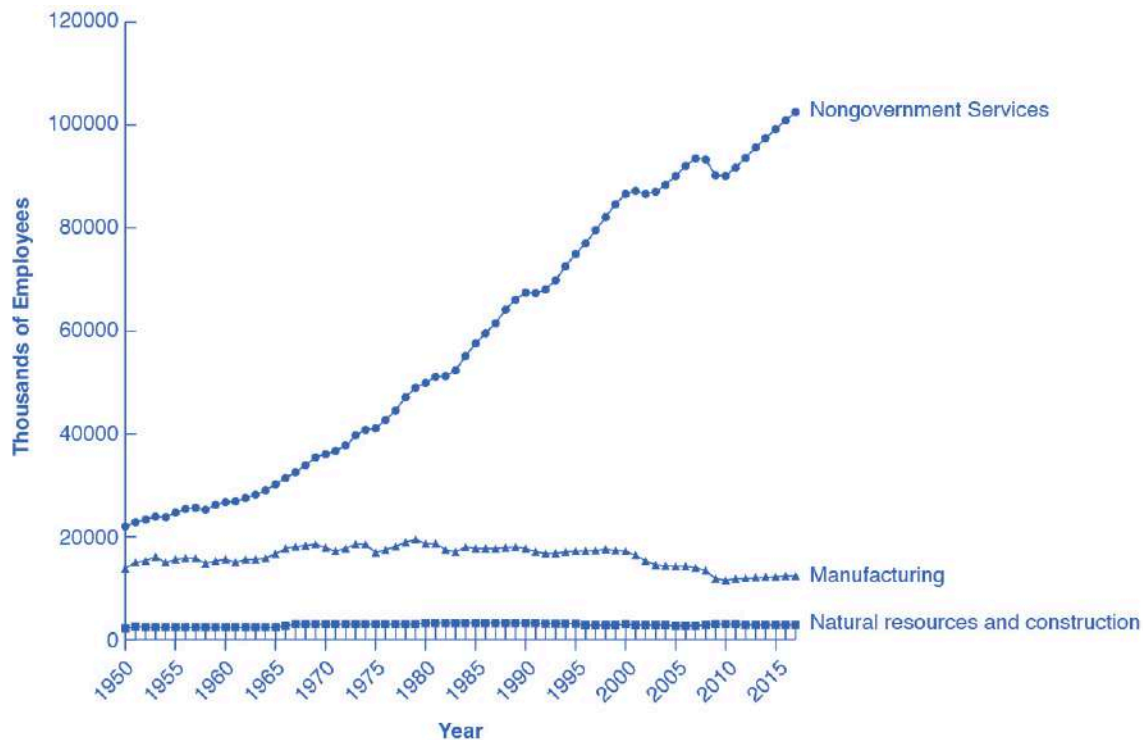


FIGURE 14.13 The Growth of Service Jobs Jobs in services have increased dramatically for more than the past 50 years. Jobs in government have increased modestly until 1990 and then declined slightly since then. Jobs in manufacturing peaked in the late 1970s and have declined more than a third since then.

A second explanation for the decline in the share of unionized workers looks at import competition. Starting in the 1960s, U.S. carmakers and steelmakers faced increasing competition from Japanese and European manufacturers. As sales of imported cars and steel rose, the number of jobs in U.S. auto manufacturing fell.

This industry is heavily unionized. Not surprisingly, membership in the United Auto Workers, which was 975,000 in 1985, had fallen to roughly 390,000 by 2015. Import competition not only decreases the employment in sectors where unions were once strong, but also decreases the bargaining power of unions in those sectors. However, as we have seen, unions that organize public-sector workers, who are not threatened by import competition, have continued to see growth.

A third possible reason for the decline in the number of union workers is that citizens often call on their elected representatives to pass laws concerning work conditions, overtime, parental leave, regulation of pensions, and other issues. Unions offered strong political support for these laws aimed at protecting workers but, in an ironic twist, the passage of those laws then made many workers feel less need for unions.

These first three possible reasons for the decline of unions are all somewhat plausible, but they have a common problem. Most other developed economies have experienced similar economic and political trends, such as the shift from manufacturing to services, globalization, and increasing government social benefits and regulation of the workplace. Clearly there are cultural differences between countries as to their acceptance of unions in the workplace. The share of the population belonging to unions in other countries is very high compared with the share in the United States. [Table 14.8](#) shows the proportion of workers in a number of the world's high-income economies who belong to unions. The United States is near the bottom, along with France and Spain. The last column shows union coverage, defined as including those workers whose wages are determined by a union negotiation even if the workers do not officially belong to the union. In the United States, union membership is almost identical to union coverage. However, in many countries, the wages of many workers who do not officially belong to a union are still determined by collective bargaining between unions and firms.

Country	Union Density: Percentage of Workers Belonging to a Union	Union Coverage: Percentage of Workers Whose Wages Are Determined by Union Bargaining
Austria	37%	99%
France	9%	95%
Germany	26%	63%
Japan	22%	23%
Netherlands	25%	82%
Spain	11.3%	81%
Sweden	82%	92%
United Kingdom	29%	35%
United States	11.1%	12.5%

TABLE 14.8 International Comparisons of Union Membership and Coverage in 2012 (Source, CIA World Factbook, retrieved from www.cia.gov)

These international differences in union membership suggest a fourth reason for the decline of union membership in the United States: perhaps U.S. laws are less friendly to the formation of unions than such laws in other countries. The close connection between union membership and a friendly legal environment is

apparent in the history of U.S. unions. The great rise in union membership in the 1930s followed the passage of the **National Labor Relations Act** of 1935, which specified that workers had a right to organize unions and that management had to give them a fair chance to do so. The U.S. government strongly encouraged forming unions during the early 1940s in the belief that unions would help to coordinate the all-out production efforts needed during World War II. However, after World War II came the passage of the Taft-Hartley Act of 1947, which gave states the power to allow workers to opt out of the union in their workplace if they so desired. This law made the legal climate less encouraging to those seeking to form unions, and union membership levels soon started declining.

The procedures for forming a union differ substantially from country to country. For example, the procedures in the United States and those in Canada are strikingly different. When a group of workers wishes to form a union in the United States, they announce this fact and set an election date when the firm's employees will vote in a secret ballot on whether to form a union. Supporters of the union lobby for a “yes” vote, and the firm's management lobbies for a “no” vote—often even hiring outside consultants for assistance in swaying workers to vote “no.” In Canada, by contrast, a union is formed when a sufficient proportion of workers (usually about 60%) signs an official card saying that they want a union. There is no separate “election date.” The management of Canadian firms is limited by law in its ability to lobby against the union. In addition, although it is illegal to discriminate and fire workers based on their union activity in the United States, the penalties are slight, making this a not so costly way of deterring union activity. In short, forming unions is easier in Canada—and in many other countries—than in the United States.

In summary, union membership in the United States is lower than in many other high-income countries, a difference that may be due to different legal environments and cultural attitudes toward unions.

LINK IT UP

Visit this [website \(http://openstax.org/l/fastfoodwages\)](http://openstax.org/l/fastfoodwages) to read more about recent protests regarding minimum wage for fast food employees.

14.4 Bilateral Monopoly

LEARNING OBJECTIVES

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

- How firms determine wages and employment when a specific labor market combines a union and a monopsony

What happens when there is market power on both sides of the labor market, in other words, when a union meets a monopsony? Economists call such a situation a **bilateral monopoly**.

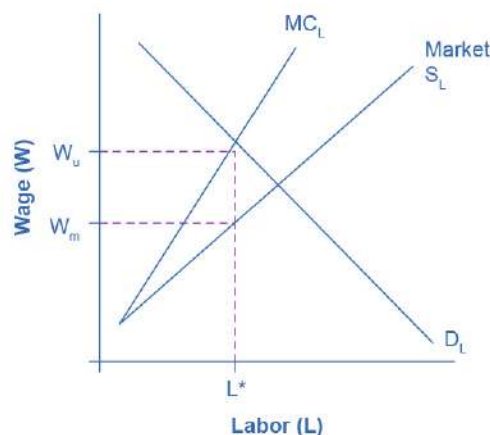


FIGURE 14.14 Bilateral Monopoly Employment, L^* , will be lower in a bilateral monopoly than in a competitive labor

market, but the equilibrium wage is indeterminate, somewhere in the range between W_u , what the union would choose, and W_m , what the monopsony would choose.

[Figure 14.14](#) is a combination of Figure 14.6 and Figure 14.11. A monopsony wants to reduce wages as well as employment, W_m and L^* in the figure. A union wants to increase wages, but at the cost of lower employment, W_u and L^* in the figure. Since both sides want to reduce employment, we can be sure that the outcome will be lower employment compared to a competitive labor market. What happens to the wage, though, is based on the monopsonist's relative bargaining power compared to the bargaining power of the union. The actual outcome is indeterminate in the graph, but it will be closer to W_u if the union has more power and closer to W_m if the monopsonist has more power.

14.5 Employment Discrimination

LEARNING OBJECTIVES

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

- Analyze earnings gaps based on race and gender
- Explain the impact of discrimination in a competitive market
- Identify U.S. public policies designed to reduce discrimination

Barriers to equitable participation in the labor market drive down economic growth. When certain populations are underrepresented, underpaid, or mistreated in a labor market or industry, the negative outcomes can effect the larger economy. For example, many science and technology fields were either unwelcoming or overtly unaccepting of women and people of color. Some major contributors to these fields overcame these challenges. Mexican-American scientist Lydia Villa-Komaroff, for example, faced overt discrimination when her college advisor told her not to pursue chemistry because women didn't "belong" in chemistry. She pursued biology instead; she developed the first instance of synthetic insulin (the chemical that people with diabetes need in order to survive) through a process that has saved million of lives and is credited with launching the entire industry of biotechnology—one of the most important in the U.S. economy. But for every Villa-Komaroff, there have been thousands of women who were prevented from making those contributions. Beyond the personal impact on those people, consider the impact on those scientific fields, our overall quality of life, and the economy itself. Economist Lisa D. Cook has quantified the costs of these innovation losses. She estimates that GDP could be as much as 4.4% higher if women and people from minority populations were fully able to participate in the science and technology innovation process.

Discrimination involves acting on the belief that members of a certain group are inferior or deserve less solely because of a factor such as race, gender, or religion. There are many types of discrimination but the focus here will be on discrimination in labor markets, which arises if workers with the same skill levels—as measured by education, experience, and expertise—receive different pay or have different job opportunities because of their race or gender. Much of the data collected and published on these topics are limited in terms of the diversity of people represented, and focus particularly on binary gender, single-race, and single-ethnicity identities. While these characterizations do not capture the diversity of Americans, the findings are important in order to understand discrimination and other practices, and to consider the impacts of policies and changes. Also, while sex and gender are different, many data sets, laws, court decisions, and media accounts use the terms interchangeably. For consistency, we will use the terminology found in the source material and government data.

Earnings Gaps by Race and Gender

A possible signal of labor market discrimination is when an employer pays one group less than another. [Figure 14.15](#) shows the average wage of Black workers as a ratio of the average wage of White workers and the average wage of female workers as a ratio of the average wage of male workers. Research by the economists Francine Blau and Laurence Kahn shows that the gap between the earnings of women and men did not move much in the 1970s, but has declined since the 1980s. Detailed analysis by economists Kerwin Kofi Charles and Patrick

Bayer show that the gap between the earnings of Black and White people diminished in the 1970s, but grew again so that current differences are as wide as they were nearly 70 years ago. In both gender and race, an earnings gap remains.

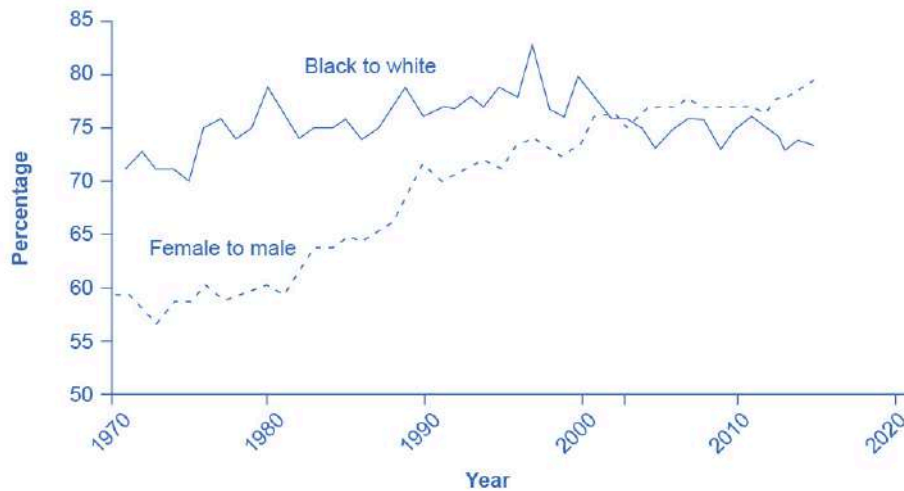


FIGURE 14.15 Wage Ratios by Sex and Race The ratio of wages for Black workers to White workers rose substantially in the late 1960s and through the 1970s. The 1990s saw a peak above 80% followed by a bumpy decline to the low 70s. The ratio of wages for female to male workers changed little through the 1970s. In both cases, a gap remains between the average wages of Black and White workers and between the average wages of female and male workers. Source: U.S. Department of Labor, Bureau of Labor Statistics.

An earnings gap between average wages, in and of itself, does not prove that discrimination is occurring in the labor market. We need to apply the same productivity characteristics to all parties (employees) involved. Gender discrimination in the labor market occurs when employers pay people of a specific gender less despite those people having comparable levels of education, experience, and expertise. (Read the Clear It Up about the sex-discrimination suit brought against Walmart.) Similarly, racial discrimination in the labor market exists when employers pay racially diverse employees less than their coworkers of the majority race despite having comparable levels of education, experience, and expertise. To bring a successful gender discrimination lawsuit, an employee must prove the employer is paying them less than an employee of a different gender who holds a similar job, with similar educational attainment, and with similar expertise. Likewise, someone who wants to sue on the grounds of racial discrimination must prove that the employer pays them less than an employee of another race who holds a similar job, with similar educational attainment, and with similar expertise.

The FRED database includes earnings data at [earnings by age, gender and race/ethnicity \(https://openstax.org/l/33501\)](https://openstax.org/l/33501).

As stated previously and as we will see below, not every instance of a wage gap or employment inequity is a product of overt discrimination on the part of individual employers. Significant overall issues in societies, such as inequitable education or housing segregation, can lead to earning gaps and limitations on economic mobility. However, these wider issues usually affect people from minority populations and/or those who have been historically underrepresented in positions of power. Economist William A. Darity Jr., whose work is discussed in more detail below, indicates that individualized employer racism still exists, but it is largely practiced in "covert and subtle forms."



CLEAR IT UP

What was the sex-discrimination case against Walmart?

In one of the largest class-action sex-discrimination cases in U.S. history, 1.2 million female employees of Walmart claimed that the company engaged in wage and promotion discrimination. In 2011, the Supreme Court threw out the case on the grounds that the group was too large and too diverse to consider the case a class action suit. Lawyers for the women regrouped and were subsequently suing in smaller groups. Part of the difficulty for the female employees is that the court said that local managers made pay and promotion decisions that were not necessarily the company's policies as a whole. Consequently, female Walmart employees in Texas argued that their new suit would challenge the management of a "discrete group of regional district and store managers." They claimed that these managers made biased pay and promotion decisions. However, in 2013, a federal district court rejected a smaller California class action suit against the company.

On other issues, Walmart made the news again in 2013 when the National Labor Relations Board found Walmart guilty of illegally penalizing and firing workers who took part in labor protests and strikes. Walmart paid \$11.7 million in back wages and compensation damages to women in Kentucky who were denied jobs due to their sex. And in 2020, a sex-based hiring discrimination lawsuit was filed by the U.S. Equal Employment Opportunity Commission (EEOC), in which the EEOC alleged that Walmart conducted a physical ability test (known as the PAT) as a requirement for applicants to be hired as order fillers at Walmart's grocery distribution centers nationwide, and that the PAT disproportionately excluded female applicants from jobs as grocery order fillers. In September 2020, Walmart and the EEOC agreed to a consent decree, which requires Walmart to cease all physical ability testing that had been used for purposes of hiring grocery distribution center order fillers. The decree also required Walmart to pay \$20 million into a settlement fund to pay lost wages to women across the country who were denied grocery order filler positions because of the testing.

Investigating the Female/Male Earnings Gap

As a result of changes in law and culture, women began to enter the paid workforce in substantial numbers in the mid- to late-twentieth century. As of February 2022, 56.0% of women aged 20 and over held jobs, while 67.6% of men aged 20 and over did. Moreover, along with entering the workforce, women began to ratchet up their education levels. In 1971, 44% of undergraduate college degrees went to women. As of the 2018–19 academic year, women earned 57% of bachelor's degrees. In 1970, women received 5.4% of the degrees from law schools and 8.4% of the degrees from medical schools. By 2017, women were receiving just over 50% of the law degrees, and by 2019, 48% of the medical degrees. There are now slightly more women than men in both law schools and medical schools. These gains in education and experience have reduced the female/male wage gap over time. However, concerns remain about the extent to which women have not yet assumed a substantial share of the positions at the top of the largest companies or in the U.S. Congress.

There are factors that can lower women's average wages. Women are likely to bear a disproportionately large share of household responsibilities. A mother of young children is more likely to drop out of the labor force for several years or work on a reduced schedule than is the father. As a result, women in their 30s and 40s are likely, on average, to have less job experience than men. In the United States, childless women with the same education and experience levels as men are typically paid comparably. However, women with families and children are typically paid about 7% to 14% less than other women of similar education and work experience. Meanwhile, married men earn about 10% to 15% more than single men with comparable education and work experience. This circumstance or practice is often referred to as the "motherhood penalty" and the "fatherhood bonus."

Another aspect of the gender pay gap relates to work that isn't actually paid, such as household chores, caring for children and other family members, and cooking. Studies have found that globally and within the United States, women undertake far more of this work than do men; even women who work full time and/or bring in

the majority of family income take on more of this unpaid work than the men in their households.

Economists study many aspects of sex- and gender-based earnings gaps, often revealing unexpected causes and impacts. For example, economists Jessica Pan, Jonathan Guryan, and Kerwin Kofi Charles analyzed decades of sociological and employment data and uncovered that the amount of sexism in the U.S. state where a woman was born is an indicator of the woman's earnings throughout her life, even if she moves away from her home state. In other words, women born in states with more pronounced sexist attitudes earn less, no matter where they live later on. Other economists showed that from 1950–2000, as women's representation increased in the workforce, jobs that became occupied by women experienced wage reductions relative to jobs being done by men—an outcome often referred to as "devaluation." The value of this research and similar investigations comes from the deeper understanding of the origins of the earnings gap, so that workers, employers, and governments can take steps to address them.

LINK IT UP

Visit this [website \(https://openstax.org/l/numberswomen\)](https://openstax.org/l/numberswomen) to read more about the persistently low numbers of women in executive roles in business and in the U.S. Congress.

Investigating the Earnings Gap Related to Race and Ethnicity

Black people experienced blatant labor market discrimination during much of the twentieth century. Until the passage of the Civil Rights Act of 1964, it was legal in many states to refuse to hire a Black worker, regardless of the credentials or experience of that worker. Moreover, Black people were often denied access to educational opportunities, which in turn meant that they had lower levels of qualifications for many jobs. At least one economic study has shown that the 1964 law is partially responsible for the narrowing of the gap in Black–White earnings in the late 1960s and into the 1970s. For example, the ratio of total earnings of Black male workers to White male workers rose from 62% in 1964 to 75.3% in 2013, according to the Bureau of Labor Statistics.

However, the earnings gap between Black and White workers has not changed as much as the earnings gap between men and women has in the last half century. The remaining racial gap seems related both to continuing differences in education levels and to the presence of discrimination. [Table 14.9](#) shows that the percentage of Black people who complete a four-year college degree remains substantially lower than the percentage of White people who complete college. According to the U.S. Census, both White and Black people have higher levels of educational attainment than Hispanic people and lower levels than Asian people. The lower average levels of education for Black workers surely explain part of the earnings gap. In fact, Black women who have the same levels of education and experience as White women receive, on average, about the same level of pay. One study shows that White and Black college graduates have identical salaries immediately after college; however, the racial wage gap widens over time, an outcome that suggests the possibility of continuing discrimination. Other researchers conducted a field experiment by responding to job advertisements with fictitious resumes using names that were either commonly associated with Black/African American people or names commonly associated with White people; they found that the White-associated names received 50 percent more callbacks for interviews. This is suggestive of discrimination in job opportunities. Further, as the following Clear It Up feature explains, there is evidence to support that discrimination in the housing market is connected to employment discrimination.

	White	Hispanic	Black	Asian
Completed four years of high school or more	93.8%	73.0%	87.2%	91.0%
Completed four years of college or more	37.6%	16.8%	23.7%	54.7%

TABLE 14.9 Educational Attainment by Race and Ethnicity for Individuals Aged 18 and Above in 2019 (Source: <https://www.census.gov/content/census/en/data/tables/2019/demo/educational-attainment/cps-detailed-tables.html>)



CLEAR IT UP

How is discrimination in the housing market connected to employment discrimination?

A recent study by the Housing and Urban Development (HUD) department found that realtors showed Black homebuyers 18 percent fewer homes compared to White homebuyers. Realtors showed Asian homebuyers 19 percent fewer properties. Additionally, Hispanic people experience more discrimination in renting apartments and undergo stiffer credit checks than White renters. In a 2012 study by the U.S. Department of Housing and Urban Development and the nonprofit Urban Institute, Hispanic testers who contacted agents about advertised rental units received information about 12 percent fewer units available and were shown seven percent fewer units than White renters. The \$9 million study, based on research in 28 metropolitan areas, concluded that blatant “door slamming” forms of discrimination are on the decline but that the discrimination that does exist is harder to detect, and as a result, more difficult to remedy. According to the *Chicago Tribune*, HUD Secretary Shaun Donovan, who served in his role from 2009-2014, told reporters, “Just because it’s taken on a hidden form doesn’t make it any less harmful. You might not be able to move into that community with the good schools.”

These practices are viewed as a continuation of **redlining**, which is the intentional and discriminatory withholding of services or products based on race or other factors. Redlining was practiced extensively by banks and other lenders who refused to issue mortgages or other loans to people from racial or ethnic minorities living in neighborhoods that were deemed “hazardous” to investment, even though the same lenders would issue loans to White people with similar economic status. Redlining has lasting effects today, demonstrated by significant divides in educational and financial opportunity in certain neighborhoods or cities.

The lower levels of education for Black workers can also be a result of discrimination—although it may be pre-labor market discrimination, rather than direct discrimination by employers in the labor market. For example, if redlining and other discrimination in housing markets causes Black families to live clustered together in certain neighborhoods and those areas have under-resourced schools, then those children will continue to have lower educational attainment than their White counterparts and, consequently, not be able to obtain the higher paying jobs that require higher levels of education. Another element to consider is that in the past, when Black people were effectively barred from many high-paying jobs, obtaining additional education could have seemed not to be worth the investment, because the educational degrees would not pay off. While the government has legally abolished discriminatory labor practices, structures and systems take a very long time to eradicate.

Competitive Markets and Discrimination

Gary Becker (1930–2014), who won the Nobel Prize in economics in 1992, was one of the first to analyze discrimination in economic terms. Becker pointed out that while competitive markets can allow some employers to practice discrimination, it can also provide profit-seeking firms with incentives not to discriminate. Given these incentives, Becker explored the question of why discrimination persists.

If a business is located in an area with a large minority population and refuses to sell to minorities, it will cut into its own profits. If some businesses run by bigoted employers refuse to pay women and/or minorities a

wage based on their productivity, then other profit-seeking employers can hire these workers. In a competitive market, if the business owners care more about the color of money than about the color of skin, they will have an incentive to make buying, selling, hiring, and promotion decisions strictly based on economic factors.

Do not underestimate the power of markets to offer at least a degree of freedom to oppressed groups. In many countries, cohesive minority population groups like Jewish people and emigrant Chinese people have managed to carve out a space for themselves through their economic activities, despite legal and social discrimination against them. Many immigrants, including those who come to the United States, have taken advantage of economic freedom to make new lives for themselves. However, history teaches that market forces alone are unlikely to eliminate discrimination. After all, discrimination against African Americans persisted in the market-oriented U.S. economy during the century between the ratification of the 13th Amendment, which abolished slavery in 1865, and the passage of the Civil Rights Act of 1964—and has continued since then, too.

Why does discrimination persist in competitive markets? Gary Becker sought to explain this persistence. Discriminatory impulses can emerge at a number of levels: among managers, among workers, and among customers. Consider the situation of a store owner or manager who is not personally prejudiced, but who has many customers who are prejudiced. If that manager treats all groups fairly, the manager may find it drives away prejudiced customers. In such a situation, a policy of nondiscrimination could reduce the firm's profits. After all, a business firm is part of society, and a firm that does not follow the societal norms is likely to suffer.

As economist William A. Darity Jr. points out, however, the "prejudiced customer" rationale falls apart when considering the many jobs that have no customer contact. Darity examined several theories regarding the persistence of employment discrimination, including rationales regarding group membership and employers' lack of information about candidates of other genders or races. Darity also directly studies and interprets others' work on discrimination in other countries, such as wage disparities between Sikh and Hindu men in India. Darity concludes that the competitive forces of the market have not been enough to overcome employment and wage discrimination, and, on their own, are unlikely to end such discrimination in the future.

LINK IT UP

Read this [article \(http://openstax.org/l/censusincome\)](http://openstax.org/l/censusincome) to learn more about wage discrimination.

Public Policies to Reduce Discrimination

A first public policy step against discrimination in the labor market is to make it illegal. For example, the Equal Pay Act of 1963 said that employers must pay men and women who do equal work the same. The Civil Rights Act of 1964 prohibits employment discrimination based on race, color, religion, sex, or national origin. The Age Discrimination in Employment Act of 1967 prohibited discrimination on the basis of age against individuals who are 40 years of age or older. The Civil Rights Act of 1991 provides monetary damages in cases of intentional employment discrimination. The Pregnancy Discrimination Act of 1978 was aimed at prohibiting discrimination against people in the workplace who are planning pregnancy, are pregnant, or are returning after pregnancy. Passing a law, however, is only part of the answer, since discrimination by prejudiced employers may be less important than broader social patterns and systems.

The 1964 Civil Rights Act created an important government organization, the Equal Employment Opportunity Commission, to investigate employment discrimination and protect workers who filed complaints against employers. Economist Phyllis Ann Wallace, who had previously worked for U.S. intelligence services, was appointed as the commission's chief of technical studies. In this role she collected and organized a massive amount of public and private sector data, as well as mentored and directed economists and other analysts in their investigations.

These laws against discrimination have reduced the gender wage gap. A 2007 Department of Labor study compared salaries of men and women who have similar educational achievement, work experience, and

occupation and found that the gender wage gap is only 5%.

In the case of the earnings gap between Black people and White people (and also between Hispanic people and White people), probably the single largest step that could be taken at this point in U.S. history to close the earnings gap would be to reduce the gap in educational attainment. Part of the answer to this issue involves finding ways to improve the performance of schools, which is a highly controversial topic in itself. In addition, the education gap is unlikely to close unless Black and Hispanic families and peer groups strengthen their culture of support for educational attainment.

Affirmative action is the name given to active efforts by government or businesses that give special rights to minorities in hiring and promotion to make up for past discrimination. Affirmative action, in its limited and not especially controversial form, means making an effort to reach out to a broader range of minority candidates for jobs. In its more aggressive and controversial form, affirmative action required government and companies to hire a specific number or percentage of minority employees. However, the U.S. Supreme Court has ruled against state affirmative action laws. Today, the government applies affirmative action policies only to federal contractors who have lost a discrimination lawsuit. The federal Equal Employment Opportunity Commission (EEOC) enforces this type of redress.

An Increasingly Diverse Workforce

Racial and ethnic diversity is on the rise in the U.S. population and workforce. As [Figure 14.16](#) shows, while the White Americans comprised 78% of the population in 2012, the U.S. Bureau of the Census projects that Whites will comprise 69% of the U.S. population by 2060. Forecasters predict that the proportion of U.S. citizens who are of Hispanic background to rise substantially. Moreover, in addition to expected changes in the population, workforce diversity is increasing as the women who entered the workforce in the 1970s and 1980s are now moving up the promotion ladders within their organizations.

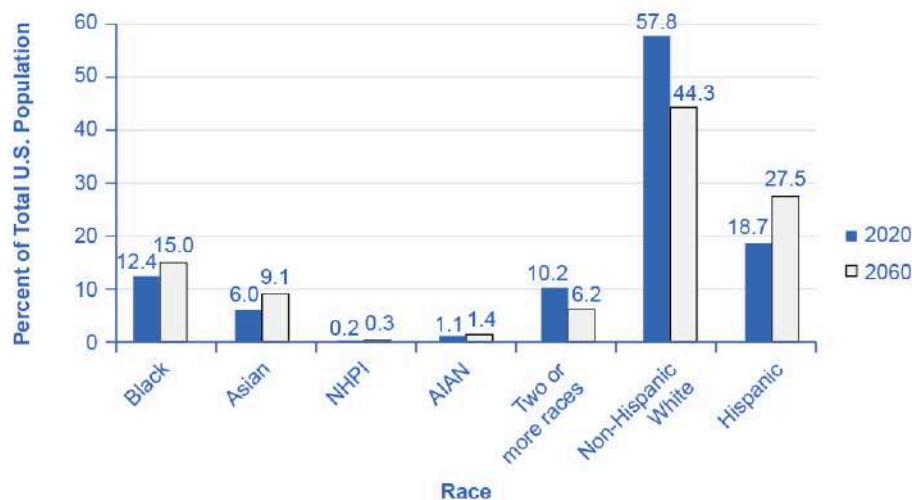


FIGURE 14.16 Projected Changes in America's Racial and Ethnic Diversity This figure shows projected changes in the ethnic makeup of the U.S. population by 2060. Note that “NHPI” stands for Native Hawaiian and Other Pacific Islander. “AIAN” stands for American Indian and Alaska Native. Source: US Department of Commerce

Regarding the future, optimists argue that the growing proportions of minority workers will break down remaining discriminatory barriers. The economy will benefit as an increasing proportion of workers from traditionally disadvantaged groups have a greater opportunity to fulfill their potential. Pessimists worry that the social tensions between different genders and between ethnic groups will rise and that workers will be less productive as a result. Anti-discrimination policy, at its best, seeks to help society move toward the more optimistic outcome.

The FRED database includes data on foreign and native born civilian [population \(https://openstax.org/l/104\)](https://openstax.org/l/104)

and [labor force \(https://openstax.org/l/32442\)](https://openstax.org/l/32442).

14.6 Immigration

Most Americans would be outraged if a law prevented them from moving to another city or another state. However, when the conversation turns to crossing national borders and is about other people arriving in the United States, laws preventing such movement often seem more reasonable. Some of the tensions over immigration stem from worries over how it might affect a country's culture, including differences in language, and patterns of family, authority, or gender relationships. Economics does not have much to say about such cultural issues. Some of the worries about immigration do, however, have to do with its effects on wages and income levels, and how it affects government taxes and spending. On those topics, economists have insights and research to offer.

Historical Patterns of Immigration

Supporters and opponents of immigration look at the same data and see different patterns. Those who express concern about immigration levels to the United States point to graphics like [Figure 14.17](#) which shows total inflows of immigrants decade by decade through the twentieth and into the twenty-first century. Clearly, the level of immigration has been high and rising in recent years, reaching and exceeding the towering levels of the early twentieth century. However, those who are less worried about immigration point out that the high immigration levels of the early twentieth century happened when total population was much lower. Since the U.S. population roughly tripled during the twentieth century, the seemingly high levels in immigration in the 1990s and 2000s look relatively smaller when they are divided by the population.

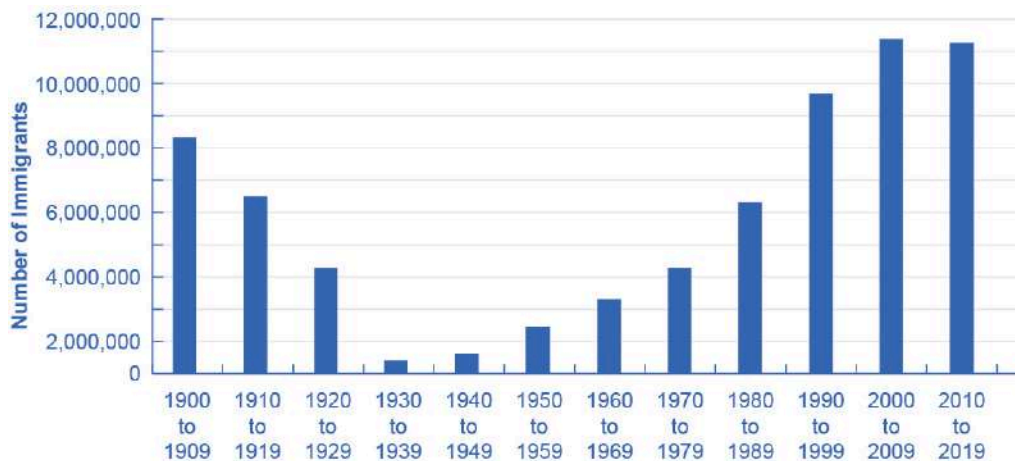


FIGURE 14.17 Immigration Since 1900 The number of immigrants in each decade declined between 1900 and the 1940s, rose sharply through 2009 and started to decline from 2010 to the present. (Source: U.S. Census)

From where have the immigrants come? Immigrants from Europe were more than 90% of the total in the first decade of the twentieth century, but less than 20% of the total by the end of the century. By the 2000s, about half of U.S. immigration came from the rest of the Americas, especially Mexico, and about a quarter came from various countries in Asia.

Economic Effects of Immigration

A surge of immigration can affect the economy in a number of different ways. In this section, we will consider how immigrants might benefit the rest of the economy, how they might affect wage levels, and how they might affect government spending at the federal and local level.

To understand the economic consequences of immigration, consider the following scenario. Imagine that the immigrants entering the United States matched the existing U.S. population in age range, education, skill levels, family size, and occupations. How would immigration of this type affect the rest of the U.S. economy?

Immigrants themselves would be much better off, because their standard of living would be higher in the United States. Immigrants would contribute to both increased production and increased consumption. Given enough time for adjustment, the range of jobs performed, income earned, taxes paid, and public services needed would not be much affected by this kind of immigration. It would be as if the population simply increased a little.

Now, consider the reality of recent immigration to the United States. Immigrants are not identical to the rest of the U.S. population. About one-third of immigrants over the age of 25 lack a high school diploma. As a result, many of the recent immigrants end up in jobs like restaurant and hotel work, lawn care, and janitorial work. This kind of immigration represents a shift to the right in the supply of unskilled labor for a number of jobs, which will lead to lower wages for these jobs. The middle- and upper-income households that purchase the services of these unskilled workers will benefit from these lower wages. However, low-skilled U.S. workers who must compete with low-skilled immigrants for jobs will tend to be negatively impacted by immigration.

The difficult policy questions about immigration are not so much about the overall gains to the rest of the economy, which seem to be real but small in the context of the U.S. economy, as they are about the disruptive effects of immigration in specific labor markets. One disruptive effect, as we noted, is that immigration weighted toward low-skill workers tends to reduce wages for domestic low-skill workers. A study by Michael S. Clune found that for each 10% rise in the number of employed immigrants with no more than a high school diploma in the labor market, high school students reduced their annual number of hours worked by 3%. The effects on wages of low-skill workers are not large—perhaps in the range of decline of about 1%. These effects are likely kept low, in part, because of the legal floor of federal and state minimum wage laws. In addition, immigrants are also thought to contribute to increased demand for local goods and services which can stimulate the local low skilled labor market. It is also possible that employers, in the face of abundant low-skill workers, may choose production processes which are more labor intensive than otherwise would have been. These various factors would explain the small negative wage effect that the native low-skill workers observed as a result of immigration.

Another potential disruptive effect is the impact on state and local government budgets. Many of the costs imposed by immigrants are costs that arise in state-run programs, like the cost of public schooling and of welfare benefits. However, many of the taxes that immigrants pay are federal taxes like income taxes and Social Security taxes. Many immigrants do not own property (such as homes and cars), so they do not pay property taxes, which are one of the main sources of state and local tax revenue. However, they do pay sales taxes, which are state and local, and the landlords of property they rent pay property taxes. According to the nonprofit Rand Corporation, the effects of immigration on taxes are generally positive at the federal level, but they are negative at the state and local levels in places where there are many low-skilled immigrants.

LINK IT UP

Visit this [website \(http://openstax.org/l/nber\)](http://openstax.org/l/nber) to obtain more context regarding immigration.

Proposals for Immigration Reform

The Congressional Jordan Commission of the 1990s proposed reducing overall levels of immigration and refocusing U.S. immigration policy to give priority to immigrants with higher skill levels. In the labor market, focusing on high-skilled immigrants would help prevent any negative effects on low-skilled workers' wages. For government budgets, higher-skilled workers find jobs more quickly, earn higher wages, and pay more in taxes. Several other immigration-friendly countries, notably Canada and Australia, have immigration systems where those with high levels of education or job skills have a much better chance of obtaining permission to immigrate. For the United States, high tech companies regularly ask for a more lenient immigration policy to admit a greater quantity of highly skilled workers under the H1B visa program.

The Obama Administration proposed the so-called “DREAM Act” legislation, which would have offered a path

to citizenship for those classified as illegal immigrants who were brought to the United States before the age of 16. Despite bipartisan support, the legislation failed to pass at the federal level. However, some state legislatures, such as California, have passed their own Dream Acts.

Between its plans for a border wall, increased deportation of undocumented immigrants, and even reductions in the number of highly skilled legal H1B immigrants, the Trump Administration had a much less positive approach to immigration. Most economists, whether conservative or liberal, believe that while immigration harms some domestic workers, the benefits to the nation exceed the costs. President Biden has been considerably more positive about immigration than his predecessor. However, given the presence of considerable disagreement within the overall population about the desirability of immigration, it is unlikely that any significant immigration reform will take place in the near future.

The FRED database includes data on the national origin of the civilian population (<https://fred.stlouisfed.org/categories/104>) (<https://fred.stlouisfed.org/categories/104>) and labor force (<https://fred.stlouisfed.org/categories/32442>) (<https://fred.stlouisfed.org/categories/32442>).



BRING IT HOME

The Increasing Value of a College Degree

The cost of college has increased dramatically in recent decades, causing many college students to take student loans to afford it. Despite this, the value of a college degree has never been higher. How can we explain this?

We can estimate the value of a bachelor's degree as the difference in lifetime earnings between the average holder of a bachelor's degree and the average high school graduate. According to a 2021 report from the Georgetown University Center on Education and the Workforce, adults with a bachelor's degree earn an average of \$2.8 million during their careers, \$1.2 million more than the median for workers with a high school diploma. College graduates also have a significantly lower unemployment rate than those with lower educational attainments.

While a college degree holder's wages have increased somewhat, the major reason for the increase in value of a bachelor's degree has been the plummeting value of a high school diploma. In the twenty-first century, the majority of jobs require at least some post-secondary education. This includes manufacturing jobs that in the past would have afforded workers a middle class income with only a high school diploma. Those jobs are increasingly scarce. This phenomenon has also no doubt contributed to the increasing inequality of income that we observe in the U.S. today. We will discuss that topic next, in Chapter 15.

Key Terms

affirmative action active efforts by government or businesses that give special rights to minorities in hiring, promotion, or access to education to make up for past discrimination

bilateral monopoly a labor market with a monopsony on the demand side and a union on the supply side

collective bargaining negotiations between unions and a firm or firms

discrimination actions based on the belief that members of a certain group or groups are in some way inferior solely because of a factor such as race, gender, or religion

first rule of labor markets an employer will never pay a worker more than the value of the worker's marginal productivity to the firm

monopsony a labor market where there is only one employer

perfectly competitive labor market a labor market where neither suppliers of labor nor demanders of labor have any market power; thus, an employer can hire all the workers they would like at the going market wage

Key Concepts and Summary

14.1 The Theory of Labor Markets

A firm demands labor because of the value of the labor's marginal productivity. For a firm operating in a perfectly competitive output market, this will be the value of the marginal product, which we define as the marginal product of labor multiplied by the firm's output price. For a firm which is not perfectly competitive, the appropriate concept is the marginal revenue product, which we define as the marginal product of labor multiplied by the firm's marginal revenue. Profit maximizing firms employ labor up to the point where the market wage is equal to the firm's demand for labor. In a competitive labor market, we determine market wage through the interaction between the market supply and market demand for labor.

14.2 Wages and Employment in an Imperfectly Competitive Labor Market

A monopsony is the sole employer in a labor market. The monopsony can pay any wage it chooses, subject to the market supply of labor. This means that if the monopsony offers too low a wage, they may not find enough workers willing to work for them. Since to obtain more workers, they must offer a higher wage, the marginal cost of additional labor is greater than the wage. To maximize profits, a monopsonist will hire workers up to the point where the marginal cost of labor equals their labor demand. This results in a lower level of employment than a competitive labor market would provide, but also a lower equilibrium wage.

14.3 Market Power on the Supply Side of Labor Markets: Unions

A labor union is an organization of workers that negotiates as a group with employers over compensation and work conditions. Union workers in the United States are paid more on average than other workers with comparable education and experience. Thus, either union workers must be more productive to match this higher pay or the higher pay will lead employers to find ways of hiring fewer union workers than they otherwise would. American union membership has been falling for decades. Some possible reasons include the shift of jobs to service industries; greater competition from globalization; the passage of worker-friendly legislation; and U.S. laws that are less favorable to organizing unions.

14.4 Bilateral Monopoly

A bilateral monopoly is a labor market with a union on the supply side and a monopsony on the demand side. Since both sides have monopoly power, the equilibrium level of employment will be lower than that for a competitive labor market, but the equilibrium wage could be higher or lower depending on which side negotiates better. The union favors a higher wage, while the monopsony favors a lower wage, but the outcome is indeterminate in the model.

14.5 Employment Discrimination

Discrimination occurs in a labor market when employers pay workers with the same economic characteristics, such as education, experience, and skill, are paid different amounts because of race, gender, religion, age, or disability status. In the United States, female workers on average earn less than male workers, and Black workers on average earn less than White workers. There is controversy over to which discrimination differences in factors like education and job experience can explain these earnings gaps. Free markets can allow discrimination to occur, but the threat of a loss of sales or a loss of productive workers can also create incentives for a firm not to discriminate. A range of public policies can be used to reduce earnings gaps between men and women or between White and other racial/ethnic groups: requiring equal pay for equal work, and attaining more equal educational outcomes.

14.6 Immigration

The recent level of U.S. immigration is at a historically high level if we measure it in absolute numbers, but not if we measure it as a share of population. The overall gains to the U.S. economy from immigration are real but relatively small. However, immigration also causes effects like slightly lower wages for low-skill workers and budget problems for certain state and local governments.

Self-Check Questions

1. [Table 14.10](#) shows levels of employment (Labor), the marginal product at each of those levels, and the price at which the firm can sell output in the perfectly competitive market where it operates.

Labor	Marginal Product of Labor	Price of the Product
1	10	\$4
2	8	\$4
3	7	\$4
4	5	\$4
5	3	\$4
6	1	\$4

TABLE 14.10

- a. What is the value of the marginal product at each level of labor?
- b. If the firm operates in a perfectly competitive labor market where the going market wage is \$12, what is the firm's profit maximizing level of employment?

2. [Table 14.11](#) shows levels of employment (Labor), the marginal product at each of those levels, and a monopoly's marginal revenue.

Labor	Marginal Product of Labor	Price of the Product
1	10	\$10
2	8	\$7
3	7	\$5
4	5	\$4
5	3	\$2
6	1	\$1

TABLE 14.11

- What is the monopoly's marginal revenue product at each level of employment?
 - If the monopoly operates in a perfectly competitive labor market where the going market wage is \$20, what is the firm's profit maximizing level of employment?
3. [Table 14.12](#) shows the quantity demanded and supplied in the labor market for driving city buses in the town of Unionville, where all the bus drivers belong to a union.

Wage Per Hour	Quantity of Workers Demanded	Quantity of Workers Supplied
\$14	12,000	6,000
\$16	10,000	7,000
\$18	8,000	8,000
\$20	6,000	9,000
\$22	4,000	10,000
\$24	2,000	11,000

TABLE 14.12

- What would the equilibrium wage and quantity be in this market if no union existed?
 - Assume that the union has enough negotiating power to raise the wage to \$4 per hour higher than it would otherwise be. Is there now excess demand or excess supply of labor?
- Do unions typically oppose new technology out of a fear that it will reduce the number of union jobs? Why or why not?
 - Compared with the share of workers in most other high-income countries, is the share of U.S. workers whose wages are determined by union bargaining higher or lower? Why or why not?
 - Are firms with a high percentage of union employees more likely to go bankrupt because of the higher wages that they pay? Why or why not?

7. Do countries with a higher percentage of unionized workers usually have less growth in productivity because of strikes and other disruptions caused by the unions? Why or why not?
8. [Table 14.13](#) shows information from the supply curve for labor for a monopsonist, that is, the wage rate required at each level of employment.

Labor	Wage
1	1
2	3
3	5
4	7
5	8
6	10

TABLE 14.13

- What is the monopsonist's marginal cost of labor at each level of employment?
 - If each unit of labor's marginal revenue product is \$13, what is the firm's profit maximizing level of employment and wage?
9. Explain in each of the following situations how market forces might give a business an incentive to act in a less discriminatory fashion.
- A local flower delivery business that had intentionally served only White customers notices that many of the local residents are Black.
 - An assembly line has traditionally only hired men, but it is having a hard time hiring sufficiently qualified workers.
 - A relationship counselor who had a strict policy of only serving straight couples notes the significant increase of LGBTQ people moving into the town.
10. Does the earnings gap between the average wages of females and the average wages of males prove labor market discrimination? Why or why not?
11. If immigration is reduced, what is the impact on the wage for low-skilled labor? Explain.

Review Questions

- What determines the demand for labor for a firm operating in a perfectly competitive output market?
- What determines the demand for labor for a firm with market power in the output market?
- What is a perfectly competitive labor market?
- What is a labor union?
- Why do employers have a natural advantage in bargaining with employees?
- What are some of the most important laws that protect employee rights?
- How does the presence of a labor union change negotiations between employers and workers?
- What is the long-term trend in American union membership?

20. Would you expect the presence of labor unions to lead to higher or lower pay for worker-members? Would you expect a higher or lower quantity of workers hired by those employers? Explain briefly.
21. What are the main causes for the recent trends in union membership rates in the United States? Why are union rates lower in the United States than in many other developed countries?
22. What is a monopsony?
23. What is the marginal cost of labor?
24. How does monopsony affect the equilibrium wage and employment levels?
25. What is a bilateral monopoly?
26. How does a bilateral monopoly affect the equilibrium wage and employment levels compared to a perfectly competitive labor market?
27. Describe how the earnings gap between men and women has evolved in recent decades.
28. Describe how the earnings gap between Black and White people has evolved in recent decades.
29. Does a gap between the average earnings of men and women, or between White and Black people, prove that employers are discriminating in the labor market? Explain briefly.
30. Will a free market tend to encourage or discourage discrimination? Explain briefly.
31. What policies, when used together with antidiscrimination laws, might help to reduce the earnings gap between men and women or between White and Black workers?
32. Describe how affirmative action is applied in the labor market.
33. What factors can explain the relatively small effect of low-skilled immigration on the wages of low-skilled workers?
34. Have levels of immigration to the United States been relatively high or low in recent years? Explain.
35. How would you expect immigration by primarily low-skill workers to affect American low-skilled workers?

Critical Thinking Questions

36. What is the marginal cost of labor for a firm that operates in a competitive labor market? How does this compare with the MCL for a monopsony?
37. Given the decline in union membership over the past 50 years, what does the theory of bilateral monopoly suggest will have happened to the equilibrium level of wages over time? Why?
38. Are unions and technological improvements complementary? Why or why not?
39. Will union membership continue to decline? Why or why not?
40. If it is not profitable to discriminate, why does discrimination persist?
41. If a company has discriminated against minorities in the past, should it be required to give priority to minority applicants today? Why or why not?
42. If the United States allows a greater quantity of highly skilled workers, what will be the impact on the average wages of highly skilled employees?
43. If all countries eliminated all barriers to immigration, would global economic growth increase? Why or why not?



FIGURE 15.1 **Occupying Wall Street** On September 17, 2011, Occupy Wall Street began in New York City's Wall Street financial district. (Credit: modification of "Occupy Wall Street Day 2 2011 Shankbone" by David Shankbone/ Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Drawing the Poverty Line
- The Poverty Trap
- The Safety Net
- Income Inequality: Measurement and Causes
- Government Policies to Reduce Income Inequality

Introduction to Poverty and Economic Inequality



BRING IT HOME

Occupy Wall Street

In September 2011, a group of protesters gathered in Zuccotti Park in New York City to decry what they perceived as increasing social and economic inequality in the United States. Calling their protest "Occupy Wall Street," they argued that the concentration of wealth among the richest 1% in the United States was both economically unsustainable and inequitable, and needed to be changed. The protest then spread to other major cities, and the Occupy movement was born.

Why were people so upset? How much wealth is concentrated among the top 1% in our society? How did they acquire so much wealth? These are very real, very important questions in the United States now, and this chapter on poverty and economic inequality will help us address the causes behind this sentiment.

The labor markets that determine the pay that workers receive do not take into account how much income a family needs for food, shelter, clothing, and health care. Market forces do not worry about what happens to families when a major local employer goes out of business. Market forces do not take time to contemplate whether those who are earning higher incomes should pay an even higher share of taxes.

However, labor markets do create considerable income inequalities. In 2020, the median American household income was \$67,521 (the median is the level where half of all families had more than that level and half had less). For family households, the median was \$86,372; for non-family households, it was \$40,464. The Census Bureau also reported that in 2020, there were 37.2 million people living in poverty, representing 11.4% of the population. Think about a family of three—perhaps a single mother with two children—attempting to pay for the basics of life on perhaps \$17,916 per year. After paying for rent, healthcare, clothing, and transportation, such a family might have \$6,000 to spend on food. Spread over 365 days, the food budget for the entire family would be about \$17 per day. To put this in perspective, most cities have restaurants where \$17 will buy you an appetizer for one.

This chapter explores how the U.S. government defines poverty, the balance between assisting the poor without discouraging work, and how federal antipoverty programs work. It also discusses income inequality—how economists measure inequality, why inequality has changed in recent decades, the range of possible government policies to reduce inequality, and the danger of a tradeoff that too great a reduction in inequality may reduce incentives for producing output.

15.1 Drawing the Poverty Line

LEARNING OBJECTIVES

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

- Explain economic inequality and how the poverty line is determined
- Analyze the U.S. poverty rate over time, noting its prevalence among different groups of citizens

Comparisons of high and low incomes raise two different issues: economic inequality and **poverty**. Poverty is measured by the number of people who fall below a certain level of income—called the **poverty line**—that defines the income one needs for a basic standard of living. **Income inequality** compares the share of the total income (or wealth) in society that different groups receive. For example, one of numerous ways to look at income inequality is to compare the share of income that the top 10% receive to the share of income that the bottom 10% receive.

In the United States, the official definition of the poverty line traces back to a single person: Mollie Orshansky. In 1963, Orshansky, who was working for the Social Security Administration, published an article called “Children of the Poor” in a highly useful and dry-as-dust publication called the *Social Security Bulletin*. Orshansky’s idea was to define a poverty line based on the cost of a healthy diet.

Her previous job had been at the U.S. Department of Agriculture, where she had worked in an agency called the Bureau of Home Economics and Human Nutrition. One task of this bureau had been to calculate how much it would cost to feed a nutritionally adequate diet to a family. Orshansky found that the average family spent one-third of its income on food. She then proposed that the poverty line be the amount one requires to buy a nutritionally adequate diet, given the size of the family, multiplied by three.

The current U.S. poverty line is essentially the same as the Orshansky poverty line, although the government adjusts the dollar amounts to represent the same buying power over time. The U.S. poverty line in 2021 ranged from \$12,880 for a single individual to \$26,500 for a household of four people.

[Figure 15.2](#) shows the U.S. **poverty rate** over time; that is, the percentage of the population below the poverty line in any given year. The poverty rate declined through the 1960s, rose in the early 1980s and early 1990s, but seems to have been slightly lower since the mid-1990s. However, in no year in the last six decades has the poverty rate been less than 10.5% of the U.S. population—that is, at best almost one American in nine is below

the poverty line. In recent years, the poverty rate peaked at 15.1% in 2010, before dropping to 10.5% in 2019. [Table 15.1](#) compares poverty rates for different groups in 2011. As you will see when we delve further into these numbers, poverty rates are relatively low for White people, for the elderly, for the well-educated, and for male-headed households. Poverty rates for females, Hispanic people, and African Americans are much higher than for White people. While Hispanic people and African Americans have a higher percentage of individuals living in poverty than others, most people in the United States living below the poverty line are White people.

LINK IT UP

Visit this [website \(http://openstax.org/l/povertyprogram\)](http://openstax.org/l/povertyprogram) for more information on U.S. poverty.

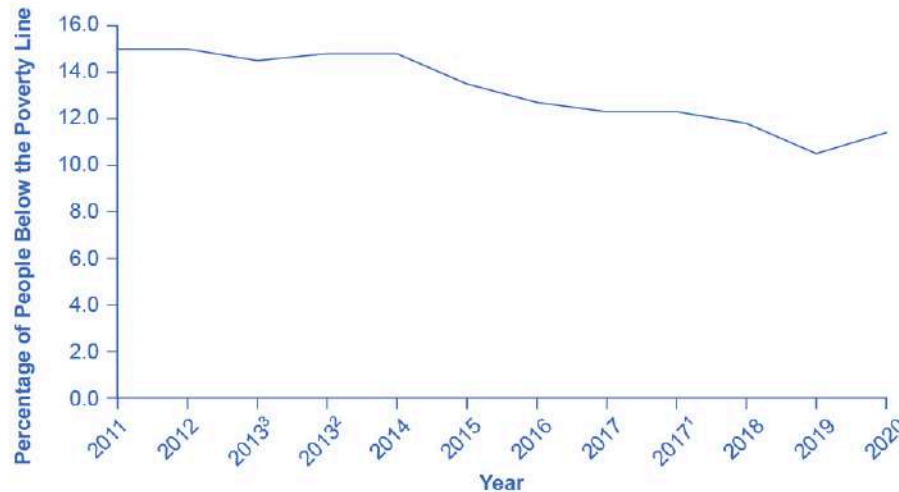


FIGURE 15.2 The U.S. Poverty Rate since 1960 The poverty rate fell dramatically during the 1960s, rose in the early 1980s and early 1990s, and, after declining in the 1990s through mid-2000s, rose to 15.1% in 2020, which is close to the 1960 levels. Between 2010 and 2019, the poverty rate declined to 10.5%, before rising to 11.4% in 2020 due to the onset of the COVID-19 pandemic in 2020. (Source: U.S. Census Bureau)

Group	Poverty Rate
Females	12.6%
Males	10.2%
White (Non-Hispanic)	8.2%
Black	19.5%
Hispanic	17.0%
Under age 18	16.1%
Ages 18–64	10.4%
Ages 65+	9.0%

TABLE 15.1 Poverty Rates by Group, 2020

The concept of a poverty line raises many tricky questions. In a vast country like the United States, should there be a national poverty line? After all, according to the Federal Register, the median household income for a family of four was \$109,113 in New Jersey and \$59,701 in Mississippi in 2017, and prices of some basic goods like housing are quite different between states. The poverty line is based on cash income, which means it does not account for government programs that provide non-cash assistance such as Medicaid (health care for low-income individuals and families) and food aid. Also, low-income families can qualify for federal housing assistance. (We will discuss these and other government aid programs in detail later in this chapter.)

Should the government adjust the poverty line to account for the value of such programs? Many economists and policymakers wonder whether we should rethink the concept of what poverty means in the twenty-first century. The following Clear It Up feature explains the poverty lines set by the World Bank for low-income countries around the world.



CLEAR IT UP

How do economists measure poverty in low-income countries?

The World Bank sets two poverty lines for low-income countries around the world. One poverty line is set at an income of \$1.90/day per person. The other is at \$3.20/day. By comparison, the U.S. 2015 poverty line of \$20,090 annually for a family of three works out to \$18.35 per person per day.

Clearly, many people around the world are far poorer than Americans, as [Table 15.2](#) shows. China and India both have more than a billion people; Nigeria is the most populous country in Africa; and Egypt is the most populous country in the Middle East. In all four of those countries, in the mid-2000s, a substantial share of the population subsisted on less than \$2/day. About half the world lives on less than \$2.50 a day, and 80 percent of the world lives on less than \$10 per day. (Of course, the cost of food, clothing, and shelter in those countries can be very different from those costs in the United States, so the \$2 and \$2.50 figures may mean greater purchasing power than they would in the United States.)

Country	Year	Percentage of Population with Income Less Than \$1.90/Day/Person	Percentage of Population with Income Less Than \$3.20/Day/Person
Brazil	2019	4.6%	9.1%
China	2017	0.5%	5.4%
Egypt	2017	3.8%	28.9%
India	2011	22.5%	61.7%
Mexico	2018	1.7%	6.5%
Nigeria	2018	39.1%	71.0%

TABLE 15.2 Poverty Lines for Low-Income Countries, mid-2000s (Source: <https://datatopics.worldbank.org/world-development-indicators/themes/poverty-and-inequality.html>)

Any poverty line will be somewhat arbitrary, and it is useful to have a poverty line whose basic definition does not change much over time. If Congress voted every few years to redefine poverty, then it would be difficult to compare rates over time. After all, would a lower poverty rate change the definition, or is it the case that people were actually better off? Government statisticians at the U.S. Census Bureau have ongoing research programs to address questions like these.

15.2 The Poverty Trap

LEARNING OBJECTIVES

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

- Explain the poverty trap, noting how government programs impact it
- Identify potential issues in government programs that seek to reduce poverty
- Calculate a budget constraint line that represents the poverty trap

Can you give people too much help, or the wrong kind of help? When people are provided with food, shelter, healthcare, income, and other necessities, assistance may reduce their incentive to work, particularly if their work is likely to offer low wages and reduce government assistance. Consider a program to fight poverty that works in this reasonable-sounding manner: the government provides assistance to the those who need it, but as the recipients earn income to support themselves, the government reduces the level of assistance it provides. With such a program, every time a person earns \$100, they lose \$100 in government support. As a result, the person experiences no net gain for working. Economists call this problem the **poverty trap**.

Consider the situation a single-parent family faces. [Figure 15.3](#) illustrates a single mother (earning \$8 an hour) with two children. First, consider the labor-leisure budget constraint that this family faces in a situation without government assistance. On the horizontal axis is hours of leisure (or time spent with family responsibilities) increasing in quantity from left to right. Also on the horizontal axis is the number of hours at paid work, going from zero hours on the right to the maximum of 2,500 hours on the left. On the vertical axis is the amount of income per year rising from low to higher amounts of income. The budget constraint line shows that at zero hours of leisure and 2,500 hours of work, the maximum amount of income is \$20,000 ($\$8 \times 2,500$ hours). At the other extreme of the budget constraint line, an individual would work zero hours, earn zero income, but enjoy 2,500 hours of leisure. At point A on the budget constraint line, by working 40 hours a week, 50 weeks a year, the utility-maximizing choice is to work a total of 2,000 hours per year and earn \$16,000.

Now suppose that a government antipoverty program guarantees every family with a single mother and two children \$18,000 in income. This is represented on the graph by a horizontal line at \$18,000. With this program, each time the mother earns \$1,000, the government will deduct \$1,000 of its support. [Table 15.3](#) shows what will happen at each combination of work and government support.

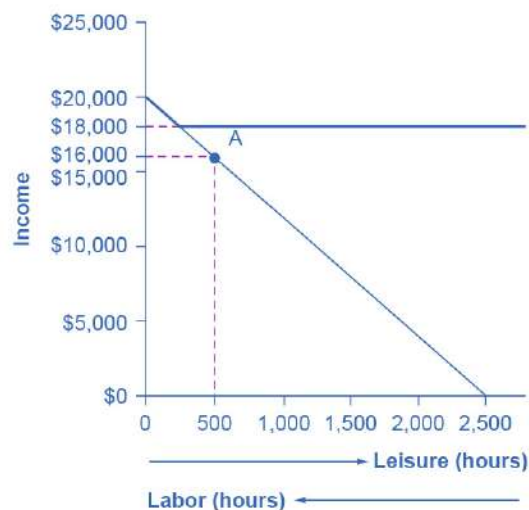


FIGURE 15.3 The Poverty Trap in Action The original choice is 500 hours of leisure, 2,000 hours of work at point A, and income of \$16,000. With a guaranteed income of \$18,000, this family would receive \$18,000 whether it provides zero hours of work or 2,000 hours of work. Only if the family provides, say, 2,300 hours of work does its income rise above the guaranteed level of \$18,000—and even then, the marginal gain to income from working many hours is small.

Amount Worked (hours)	Total Earnings	Government Support	Total Income
0	0	\$18,000	\$18,000
500	\$4,000	\$14,000	\$18,000
1,000	\$8,000	\$10,000	\$18,000
1,500	\$12,000	\$6,000	\$18,000
2,000	\$16,000	\$2,000	\$18,000
2,500	\$20,000	0	\$20,000

TABLE 15.3 Total Income at Various Combinations of Work and Support

The new budget line, with the antipoverty program in place, is the horizontal and heavy line that is flat at \$18,000. If the mother does not work at all, she receives \$18,000, all from the government. If she works full time, giving up 40 hours per week with her children, she still ends up with \$18,000 at the end of the year. Only if she works 2,300 hours in the year—which is an average of 44 hours per week for 50 weeks a year—does household income rise to \$18,400. Even in this case, all of her year’s work means that household income rises by only \$400 over the income she would receive if she did not work at all. She would need to work 50 hours a week to reach \$20,800.

The poverty trap is even stronger than this simplified example shows, because a working mother will have extra expenses like clothing, transportation, and child care that a nonworking mother will not face, making the economic gains from working even smaller. Moreover, those who do not work fail to build up job experience and contacts, which makes working in the future even less likely.

To reduce the poverty trap the government could design an antipoverty program so that, instead of reducing government payments by \$1 for every \$1 earned, the government would reduce payments by some smaller amount instead. Imposing requirements for work as a condition of receiving benefits and setting a time limit on benefits can also reduce the harshness of the poverty trap.

[Figure 15.4](#) illustrates a government program that guarantees \$18,000 in income, even for those who do not work at all, but then reduces this amount by 50 cents for each \$1 earned. The new, higher budget line in [Figure 15.4](#) shows that, with this program, additional hours of work will bring some economic gain. Because of the reduction in government income when an individual works, an individual earning \$8.00 will really net only \$4.00 per hour. The vertical intercept of this higher budget constraint line is at \$28,000 ($\$18,000 + 2,500 \text{ hours} \times \$4.00 = \$28,000$). The horizontal intercept is at the point on the graph where \$18,000 and 2500 hours of leisure is set. [Table 15.4](#) shows the total income differences with various choices of labor and leisure.

However, this type of program raises other issues. First, even if it does not eliminate the incentive to work by reducing government payments by \$1 for every \$1 earned, enacting such a program may still reduce the incentive to work. At least some people who would be working 2,000 hours each year without this program might decide to work fewer hours but still end up with more income—that is, their choice on the new budget line would be like S, above and to the right of the original choice P. Of course, others may choose a point like R, which involves the same amount of work as P, or even a point to the left of R that involves more work.

The second major issue is that when the government phases out its support payments more slowly, the antipoverty program costs more money. Still, it may be preferable in the long run to spend more money on a program that retains a greater incentive to work, rather than spending less money on a program that nearly eliminates any gains from working.

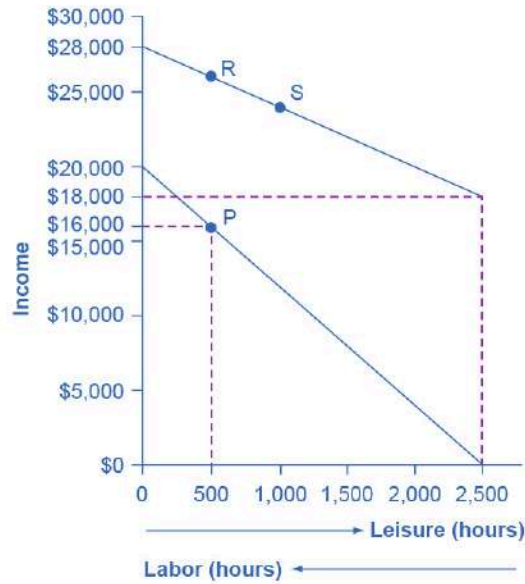


FIGURE 15.4 Loosening the Poverty Trap: Reducing Government Assistance by 50 Cents for Every \$1 Earned On the original labor-leisure opportunity set, the lower, downward-sloping budget set, the preferred choice P is 500 hours of leisure and \$16,000 of income. Then, the government created an antipoverty program that guarantees \$18,000 in income even to those who work zero hours, shown by the horizontal dashed line. In addition, every \$1 earned means phasing out 50 cents of benefits at \$18,000. This program leads to the higher budget set, which the diagram shows. The hope is that this program will provide incentives to work the same or more hours, despite receiving income assistance. However, it is possible that the recipients will choose a point on the new budget set like S, with less work, more leisure, and greater income, or a point like R, with the same work and greater income.

Amount Worked (hours)	Total Earnings	Government Support	Total Income
0	0	\$18,000	\$18,000
500	\$4,000	\$16,000	\$20,000
1,000	\$8,000	\$14,000	\$22,000
1,500	\$12,000	\$12,000	\$24,000
2,000	\$16,000	\$10,000	\$26,000
2,500	\$20,000	\$8,000	\$28,000

TABLE 15.4 The Labor-Leisure Tradeoff with Assistance Reduced by 50 Cents for Every Dollar Earned

The next module will consider a variety of government support programs focused specifically on people experiencing poverty, including welfare, SNAP (Supplemental Nutrition Assistance Program), Medicaid, and the earned income tax credit (EITC). Although these programs vary from state to state, it is generally a true statement that in many states from the 1960s into the 1980s, if poor people worked, their level of income barely rose—or did not rise at all—after factoring in the reduction in government support payments. The following Work It Out feature shows how this happens.

15.3 The Safety Net

LEARNING OBJECTIVES

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

- Identify the antipoverty government programs that comprise the safety net
- Explain the safety net programs' primary goals and how these programs have changed over time
- Discuss the complexities of these safety net programs and why they can be controversial

The U.S. government has implemented a number of programs to assist those below the poverty line and those who have incomes just above the poverty line. Such programs are called the **safety net**, to recognize that they offer some protection for those who find themselves without jobs or income.

Temporary Assistance for Needy Families

From the Great Depression until 1996, the United States' most visible antipoverty program was Aid to Families with Dependent Children (AFDC), which provided cash payments to mothers with children who were below the poverty line. Many just called this program “welfare.” In 1996, Congress passed and President Bill Clinton signed into law the Personal Responsibility and Work Opportunity Reconciliation Act, more commonly called the “welfare reform act.” The new law replaced AFDC with Temporary Assistance for Needy Families (TANF).

LINK IT UP

Visit this [website \(http://openstax.org/l/Clinton_speech\)](http://openstax.org/l/Clinton_speech) to watch a video of President Bill Clinton's Welfare Reform speech.

TANF brought several dramatic changes in how welfare operated. Under the old AFDC program, states set the level of welfare benefits that they would pay to people experiencing poverty, and the federal government guaranteed it would chip in some of the money as well. The federal government's welfare spending would rise or fall depending on the number of people in need, and on how each state set its own welfare contribution.

Under TANF, however, the federal government gives a fixed amount of money to each state. The state can then use the money for almost any program with an antipoverty component: for example, the state might use the money to give funds to families with low income, or to reduce teenage pregnancy, or even to raise the high school graduation rate. However, the federal government imposed two key requirements. First, if states are to keep receiving the TANF grants, they must impose work requirements so that most of those receiving TANF benefits are working (or attending school). Second, no one can receive TANF benefits with federal money for more than a total of five years over their lifetime. The old AFDC program had no such work requirements or time limits.

TANF attempts to avoid the poverty trap by requiring that welfare recipients work and by limiting the length of time they can receive benefits. In its first few years, the program was quite successful. The number of families receiving payments in 1995, the last year of AFDC, was 4.8 million. November 2020, according to the Congressional Research Service, the number of families receiving payments under TANF was 1.0 million—a decline of nearly 80%.

TANF benefits to poor families vary considerably across states. For example, again according to the Congressional Research Service, in July 2020 the highest monthly payment in New Hampshire to a single mother with one child was \$862, while in Mississippi the highest monthly payment to that family was \$146. In part, these payments reflect differences in states' cost of living. As reported by the Department of Health and Human Services, in 1995 total spending on TANF was approximately \$19 billion. Spending increased yearly through 2001, then it was roughly flat at approximately \$26 billion until 2005, then it increased again through 2010, where it peaked at nearly \$35 billion. It then decreased again to around \$30 billion in 2020. When you take into account the effects of inflation, the decline is even greater. Moreover, there seemed little evidence that families were suffering a reduced standard of living as a result of TANF—although, on the other side, there was

not much evidence that families had greatly improved their total levels of income, either.

The Earned Income Tax Credit (EITC)

The **earned income tax credit (EITC)**, first passed in 1975, is a method of assisting the working poor through the tax system. The EITC is one of the largest assistance program for low-income groups, and as of December 2021, about 25 million eligible workers and families received about \$60 billion in EITC. For the 2021 tax year, the earned income credit ranges from \$1,502 to \$6,728 depending on tax-filing status, income, and number of children. The average amount of EITC received nationwide was about \$2,411. In 2021, for example, a single parent with two children would have received a tax credit of \$5,980 up to a modest income level. The amount of the tax break increases with the amount of income earned, up to a point. The earned income tax credit has often been popular with both economists and the general public because of the way it effectively increases the payment received for work.

What about the danger of the poverty trap that every additional \$1 earned will reduce government support payments by close to \$1? To minimize this problem, the earned income tax credit is phased out slowly. For example, according to the Tax Policy Center, for a single-parent family with two children in 2013, the credit is not reduced at all (but neither is it increased) as earnings rise from \$13,430 to \$17,530. Then, for every \$1 earned above \$17,530, the amount received from the credit is reduced by 21.06 cents, until the credit phases out completely at an income level of \$46,227.

Figure 15.5 illustrates that the earned income tax credits, child tax credits, and the TANF program all cost the federal government money—either in direct outlays or in loss of tax revenues. CTC stands for the government tax cuts for the child tax credit.

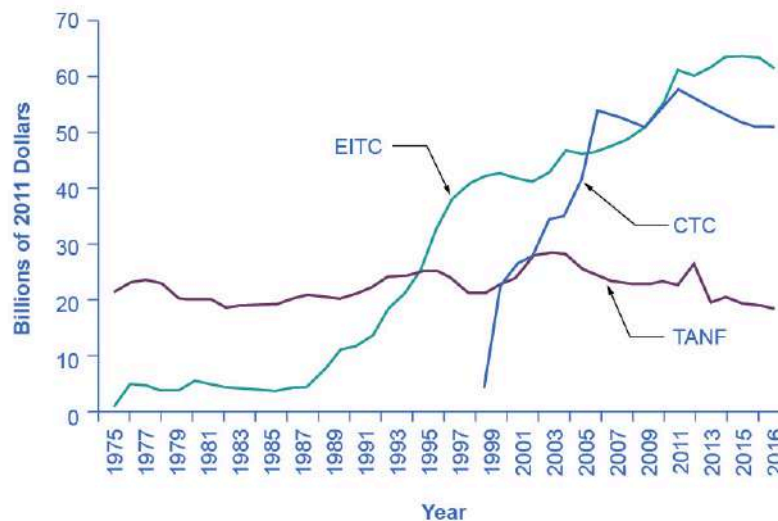


FIGURE 15.5 Real Federal Spending on CTC, EITC, and TANF, 1975–2016 EITC increased from under \$10 billion in the late 1980s to almost \$42 billion in 2000 and to over \$61 billion in 2016, far exceeding estimated 2016 outlays in the CTC (Child Tax Credits) and TANF of over \$25 billion and \$18 billion, respectively. (Source: Office of Management and Budget)

In recent years, the EITC has become a hugely expensive government program for providing income assistance to people below or near the poverty line, costing about \$60 billion in 2021. In that year, the EITC provided benefits to about 25 million families and individuals and, on average, is worth about \$2,411 per family (with children), according to the Tax Policy Center. One reason that the TANF law worked as well as it did is that the government greatly expanded EITC in the late 1980s and again in the early 1990s, which increased the returns to work for low-income Americans.

Supplemental Nutrition Assistance Program (SNAP)

Often called “food stamps,” **Supplemental Nutrition Assistance Program (SNAP)** is a federally funded program, started in 1964, in which each month people receive a card like a debit card that they can use to buy food. The amount of food aid for which a household is eligible varies by income, number of children, and other factors but, in general, households are expected to spend about 30% of their own net income on food, and if 30% of their net income is not enough to purchase a nutritionally adequate diet, then those households are eligible for SNAP.

SNAP can contribute to the poverty trap. For every \$100 earned, the government assumes that a family can spend \$30 more for food, and thus reduces its eligibility for food aid by \$30. This decreased benefit is not a complete disincentive to work—but combined with how other programs reduce benefits as income increases, it adds to the problem. SNAP, however, does try to address the poverty trap with its own set of work requirements and time limits.

Why give debit cards and not just cash? Part of the political support for SNAP comes from a belief that since recipients must spend the cards on food, they cannot “waste” them on other forms of consumption. From an economic point of view, however, the belief that cards must increase spending on food seems wrong-headed. After all, say that a family is spending \$2,500 per year on food, and then it starts receiving \$1,000 per year in SNAP aid. The family might react by spending \$3,500 per year on food (income plus aid), or it might react by continuing to spend \$2,500 per year on food, but use the \$1,000 in food aid to free up \$1,000 that it can now spend on other goods. Thus, it is reasonable to think of SNAP cards as an alternative method, along with TANF and the earned income tax credit, of transferring income to those working but still experiencing poverty.

Anyone eligible for TANF is also eligible for SNAP, although states can expand eligibility for food aid if they wish to do so. In some states, where TANF welfare spending is relatively low, a poor family may receive more in support from SNAP than from TANF. In 2021, about 41.5 million people received food aid with total benefits of just over \$108 billion, which is an average monthly benefit of about \$287 per person per month. SNAP participation increased by 70% between 2007 and 2011, from 26.6 million participants to 45 million. According to the Congressional Budget Office, the 2008-2009 Great Recession and rising food prices caused this dramatic rise in participation. Likewise, between 2019 and 2021, the number of participants in SNAP increased by 5.8 million, the amount per person increased by 67%, and total benefits nearly doubled as a consequence of the sharp recession due to the onset of the COVID-19 pandemic in early 2020.

The federal government deploys a range of income security programs that it funds through departments such as Health and Human Services, Agriculture, and Housing and Urban Development (HUD) (see [Figure 15.6](#)). According to the Office of Management and Budget, collectively, these three departments provided an estimated \$62 billion of aid through programs such as supplemental feeding programs for women and children, subsidized housing, and energy assistance. The federal government also transfers funds to individual states through special grant programs.

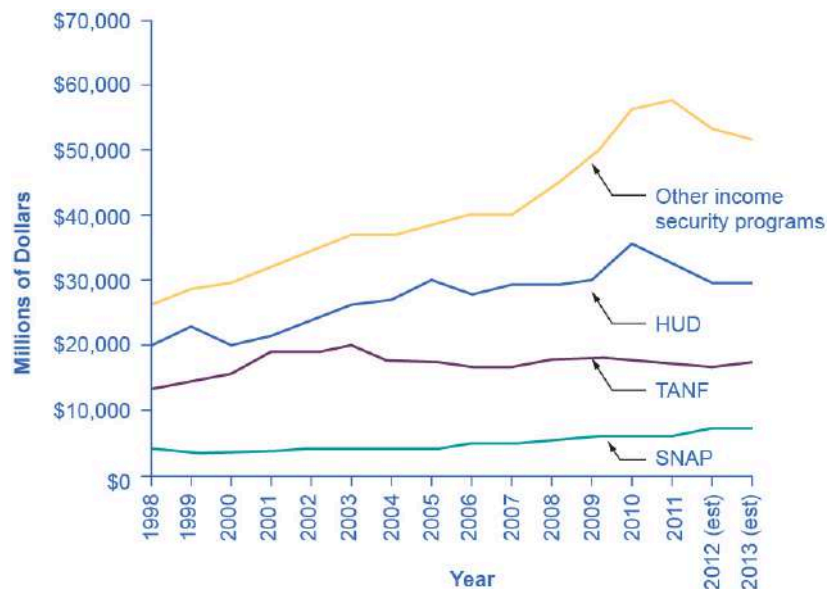


FIGURE 15.6 Expenditure Comparison of TANF, SNAP, HUD, and Other Income Security Programs, 1988–2013 (est.) Total expenditures on income security continued to rise between 1988 and 2010, while payments for TANF have increased from \$13 billion in 1998 to an estimated \$17.3 billion in 2013. SNAP has seen relatively small increments. These two programs comprise a relatively small portion of the estimated \$106 billion dedicated to income security in 2013. Note that other programs and housing programs increased dramatically during the 2008 and 2010 time periods. (Source: Table 12.3 Section 600 Income Security, <https://www.whitehouse.gov/sites/default/files/omb/budget/fy2013/assets/hist.pdf>)

The safety net includes a number of other programs: government-subsidized school lunches and breakfasts for children from low-income families; the Special Supplemental Food Program for Women, Infants and Children (WIC), which provides food assistance for pregnant women and newborns; the Low Income Home Energy Assistance Program, which provides help with home heating bills; housing assistance, which helps pay the rent; and Supplemental Security Income, which provides cash support for people with disabilities and elderly people experiencing poverty.

Medicaid

Congress created **Medicaid** in 1965. This is a joint health insurance program between both the states and the federal government. The federal government helps fund Medicaid, but each state is responsible for administering the program, determining the level of benefits, and determining eligibility. It provides medical insurance for certain people with low incomes, including those below the poverty line, with a focus on families with children, the elderly, and people with disabilities. About one-third of Medicaid spending is for low-income mothers with children. While an increasing share of the program funding in recent years has gone to pay for nursing home costs for older people who cannot afford to pay for housing. The program ensures that participants receive a basic level of benefits, but because each state sets eligibility requirements and provides varying levels of service, the program differs from state to state.

In the past, a common problem has been that many low-paying jobs pay enough to a breadwinner so that a family could lose its eligibility for Medicaid, yet the job does not offer health insurance benefits. A parent considering such a job might choose not to work rather than lose health insurance for their children. In this way, health insurance can become a part of the poverty trap. Many states recognized this problem in the 1980s and 1990s and expanded their Medicaid coverage to include people earning up to 135% or even 185% of the poverty line. Some states also guaranteed that children would not lose coverage if their parents worked.

These expanded guarantees cost the government money, of course, but they also helped to encourage those on welfare to enter the labor force. As of 2014, approximately 69.7 million people participated in Medicaid. Of

those enrolled, almost half are children. Healthcare expenditures, however, are highest for the elderly population, which comprises approximately 25% of participants. As [Figure 15.7](#) (a) indicates, the largest number of households that enroll in Medicaid are those with children. Lower-income adults are the next largest group enrolled in Medicaid at 28%. People who are blind or have a disability account for 16% of those enrolled, and seniors are 9% of those enrolled. [Figure 15.7](#) (b) shows how much actual Medicaid dollars the government spends for each group. Out of total Medicaid spending, the government spends more on seniors (20%) and people who are blind or have a disability (44%). Thus, 64% of all Medicaid spending goes to seniors, those who are blind, and people with disabilities. Children receive 21% of all Medicaid spending, followed by adults at 15%.

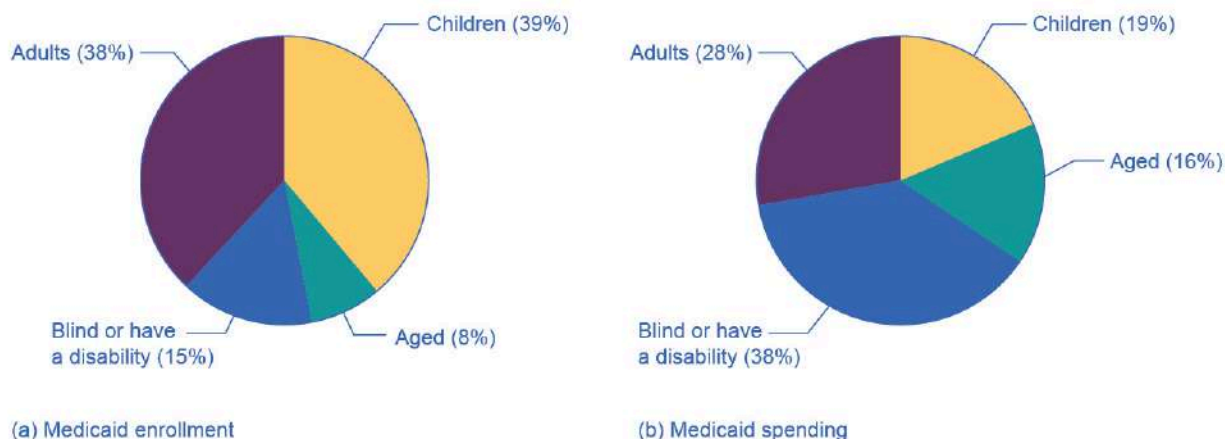


FIGURE 15.7 Medicaid Enrollment and Spending Part (a) shows the Medicaid enrollment by different populations, with children comprising the largest percentage at 47%, followed by adults at 28%, and those who are blind or have a disability at 16%. Part (b) shows that Medicaid spending is principally for those who are blind or have a disability, followed by the elderly. Although children are the largest population that Medicaid covers, expenditures on children are only at 19%.

15.4 Income Inequality: Measurement and Causes

LEARNING OBJECTIVES

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

- Explain the distribution of income, and analyze the sources of income inequality in a market economy
- Measure income distribution in quintiles
- Calculate and graph a Lorenz curve
- Show income inequality through demand and supply diagrams

Poverty levels can be subjective based on the overall income levels of a country. Typically a government measures poverty based on a percentage of the median income. Income inequality, however, has to do with the distribution of that income, in terms of which group receives the most or the least income. Income inequality involves comparing those with high incomes, middle incomes, and low incomes—not just looking at those below or near the poverty line. In turn, measuring income inequality means dividing the population into various groups and then comparing the groups, a task that we can carry out in several ways, as the next Clear It Up feature shows.



CLEAR IT UP

How do you separate poverty and income inequality?

Poverty can change even when inequality does not move at all. Imagine a situation in which income for everyone in the population declines by 10%. Poverty would rise, since a greater share of the population would now fall below

the poverty line. However, inequality would be the same, because everyone suffered the same proportional loss. Conversely, a general rise in income levels over time would keep inequality the same, but reduce poverty.

It is also possible for income inequality to change without affecting the poverty rate. Imagine a situation in which a large number of people who already have high incomes increase their incomes by even more. Inequality would rise as a result—but the number of people below the poverty line would remain unchanged.

Why did inequality of household income increase in the United States in recent decades? A trend toward greater income inequality has occurred in many countries around the world, although the effect has been more powerful in the U.S. economy. Economists have focused their explanations for the increasing inequality on two factors that changed more or less continually from the 1970s into the 2000s. One set of explanations focuses on the changing shape of American households. The other focuses on greater inequality of wages, what some economists call “winner take all” labor markets. We will begin with how we measure inequality, and then consider the explanations for growing inequality in the United States.

Measuring Income Distribution by Quintiles

One common way of measuring income inequality is to rank all households by income, from lowest to highest, and then to divide all households into five groups with equal numbers of people, known as **quintiles**. This calculation allows for measuring the distribution of income among the five groups compared to the total. The first quintile is the lowest fifth or 20%, the second quintile is the next lowest, and so on. We can measure income inequality by comparing what share of the total income each quintile earns.

U.S. income distribution by quintile appears in [Table 15.5](#). In 2020, for example, the bottom quintile of the income distribution received 3.2% of income; the second quintile received 8.1%; the third quintile, 14.0%; the fourth quintile, 22.6%; and the top quintile, 52.2%. The final column of [Table 15.5](#) shows what share of income went to households in the top 5% of the income distribution: 23.0% in 2020. Over time, from the late 1960s to the early 1980s, the top fifth of the income distribution typically received between about 43% to 44% of all income. The share of income that the top fifth received then begins to rise. Census Bureau researchers trace, much of this increase in the share of income going to the top fifth to an increase in the share of income going to the top 5%. The quintile measure shows how income inequality has increased in recent decades.

Year	Lowest Quintile	Second Quintile	Third Quintile	Fourth Quintile	Highest Quintile	Top 5%
1967	4.0	10.8	17.3	24.2	43.6	17.2
1970	4.1	10.8	17.4	24.5	43.3	16.6
1975	4.3	10.4	17.0	24.7	43.6	16.5
1980	4.2	10.2	16.8	24.7	44.1	16.5
1985	3.9	9.8	16.2	24.4	45.6	17.6
1990	3.8	9.6	15.9	24.0	46.6	18.5
1995	3.7	9.1	15.2	23.3	48.7	21.0
2000	3.6	8.9	14.8	23.0	49.8	22.1

TABLE 15.5 Share of Aggregate Income Received by Each Fifth and Top 5% of Households, 1967–2020

(Source: U.S. Census Bureau, <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html>, Table H-1, All Races.)

Year	Lowest Quintile	Second Quintile	Third Quintile	Fourth Quintile	Highest Quintile	Top 5%
2005	3.4	8.6	14.6	23.0	50.4	22.2
2010	3.3	8.5	14.6	23.4	50.3	21.3
2015	3.1	8.2	14.3	23.2	51.1	22.1
2020	3.0	8.1	14.0	22.6	52.2	23.0

TABLE 15.5 Share of Aggregate Income Received by Each Fifth and Top 5% of Households, 1967–2020
(Source: U.S. Census Bureau, <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html>, Table H-1, All Races.)

It can also be useful to divide the income distribution in ways other than quintiles; for example, into tenths or even into percentiles (that is, hundredths). A more detailed breakdown can provide additional insights. For example, the last column of [Table 15.5](#) shows the income received by the top 5% of the income distribution. Between 1980 and 2020, the share of income going to the top 5% increased by 6.5 percentage points (from 16.5% in 1980 to 23.0% in 2020). From 1980 to 2020 the share of income going to the top quintile increased by 8.1 percentage points (from 44.1% in 1980 to 52.2% in 2013). Thus, the top 20% of householders (the fifth quintile) received over half (51%) of all the income in the United States in 2020.

Lorenz Curve

We can present the data on income inequality in various ways. For example, you could draw a bar graph that showed the share of income going to each fifth of the income distribution. [Figure 15.8](#) presents an alternative way of showing inequality data in a **Lorenz curve**. This curve shows the cumulative share of population on the horizontal axis and the cumulative percentage of total income received on the vertical axis.

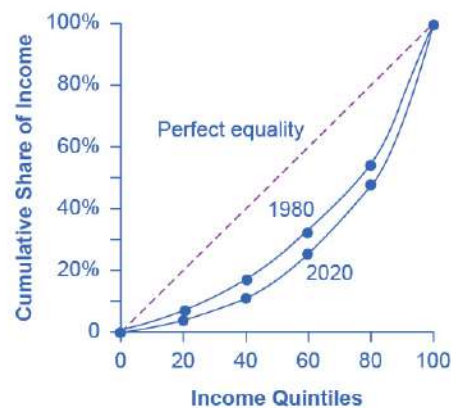


FIGURE 15.8 The Lorenz Curve A Lorenz curve graphs the cumulative shares of income received by everyone up to a certain quintile. The income distribution in 1980 was closer to the perfect equality line than the income distribution in 2020—that is, the U.S. income distribution became more unequal over time.

Every Lorenz curve diagram begins with a line sloping up at a 45-degree angle. We show it as a dashed line in [Figure 15.8](#). The points along this line show what perfect equality of the income distribution looks like. It would mean, for example, that the bottom 20% of the income distribution receives 20% of the total income, the bottom 40% gets 40% of total income, and so on. The other lines reflect actual U.S. data on inequality for 1980 and 2020.

The trick in graphing a Lorenz curve is that you must change the shares of income for each specific quintile, which we show in the first and third columns of numbers in [Table 15.6](#), into cumulative income, which we

show in the second and fourth columns of numbers. For example, the bottom 40% of the cumulative income distribution will be the sum of the first and second quintiles; the bottom 60% of the cumulative income distribution will be the sum of the first, second, and third quintiles, and so on. The final entry in the cumulative income column needs to be 100%, because by definition, 100% of the population receives 100% of the income.

Income Category	Share of Income in 1980 (%)	Cumulative Share of Income in 1980 (%)	Share of Income in 2020 (%)	Cumulative Share of Income in 2020 (%)
First quintile	4.2	4.2	3.0	3.0
Second quintile	10.2	14.4	8.1	11.1
Third quintile	16.8	31.2	14.0	25.1
Fourth quintile	24.7	55.9	22.6	47.7
Fifth quintile	44.1	100.0	52.2	100.0

TABLE 15.6 Calculating the Lorenz Curve

In a Lorenz curve diagram, a more unequal distribution of income will loop farther down and away from the 45-degree line, while a more equal distribution of income will move the line closer to the 45-degree line. [Figure 15.8](#) illustrates the greater inequality of the U.S. income distribution between 1980 and 2020 because the Lorenz curve for 2020 is farther from the 45-degree line than the Lorenz curve for 1980. The Lorenz curve is a useful way of presenting the quintile data that provides an image of all the quintile data at once. The next Clear It Up feature shows how income inequality differs in various countries compared to the United States.



CLEAR IT UP

How does economic inequality vary around the world?

The U.S. economy has a relatively high degree of income inequality by global standards. As [Table 15.7](#) shows, based on a variety of national surveys for a selection of years in the second decade of this century, the U.S. economy has greater inequality than Germany (along with most Western European countries). The region of the world with the highest level of income inequality is Latin America, illustrated in the numbers for Brazil and Mexico. The level of inequality in the United States is higher than in some of the low-income countries of the world, like India and Nigeria, as well as in some middle-income countries, like China and Russia.

Country	Survey Year	First Quintile	Second Quintile	Third Quintile	Fourth Quintile	Fifth Quintile
United States	2020	3.0%	8.1%	14.0%	22.6%	52.2%
Germany	2016	7.6%	12.8%	17.1%	22.8%	39.6%

TABLE 15.7 Income Distribution in Select Countries (Source: U.S. data from U.S. Census Bureau Table H-1. Other data from The World Bank Poverty and Inequality Data Base, <https://datatopics.worldbank.org/world-development->

Country	Survey Year	First Quintile	Second Quintile	Third Quintile	Fourth Quintile	Fifth Quintile
Brazil	2019	3.1%	7.4%	12.3%	19.4%	57.8%
Mexico	2018	5.4%	9.5%	13.5%	20.0%	51.7%
China	2016	6.5%	10.7%	15.3%	22.2%	45.3%
India	2011	8.1%	11.7%	15.2%	20.5%	44.4%
Russia	2018	7.1%	11.2%	15.2%	21.4%	45.1%
Nigeria	2018	7.1%	11.6%	16.2%	22.7%	42.4%

TABLE 15.7 Income Distribution in Select Countries (Source: U.S. data from U.S. Census Bureau Table H-1. Other data from The World Bank Poverty and Inequality Data Base, <https://datatopics.worldbank.org/world-development-indicators/themes/poverty-and-inequality.html>)

LINK IT UP

Visit this [website \(http://openstax.org/l/inequality/\)](http://openstax.org/l/inequality/) to watch a video of wealth inequality across the world.

Causes of Growing Inequality: The Changing Composition of American Households

In 1970, 41% of married women were in the labor force, but by 2019, according to the Bureau of Labor Statistics, 58.6% of married women were in the labor force. One result of this trend is that more households have two earners. Moreover, it has become more common for one high earner to marry another high earner. A few decades ago, the common pattern featured a man with relatively high earnings, such as an executive or a doctor, marrying a woman who did not earn as much, like a secretary or a nurse. Often, the woman would leave paid employment, at least for a few years, to raise a family. However, now doctors are marrying doctors and executives are marrying executives, and mothers with high-powered careers are often returning to work while their children are quite young. This pattern of households with two high earners tends to increase the proportion of high-earning households.

According to data in the National Journal, even as two-earner couples have increased, so have single-parent households. Of all U.S. families, in 2021, about 23% were headed by single mothers. The poverty rate among single-parent households tends to be relatively high.

These changes in family structure, including the growth of single-parent families who tend to be at the lower end of the income distribution, and the growth of two-career high-earner couples near the top end of the income distribution, account for roughly half of the rise in income inequality across households in recent decades.

LINK IT UP

Visit this [website \(http://openstax.org/l/US_wealth\)](http://openstax.org/l/US_wealth) to watch a video that illustrates the distribution of wealth in the United States.

Causes of Growing Inequality: A Shift in the Distribution of Wages

Another factor behind the rise in U.S. income inequality is that earnings have become less equal since the late 1970s. In particular, the earnings of high-skilled labor relative to low-skilled labor have increased. Winner-take-all labor markets result from changes in technology, which have increased global demand for

“stars,”—whether the best CEO, doctor, basketball player, or actor. This global demand pushes salaries far above productivity differences associated with educational differences. One way to measure this change is to take workers' earnings with at least a four-year college bachelor's degree (including those who went on and completed an advanced degree) and divide them by workers' earnings with only a high school degree. The result is that those in the 25–34 age bracket with college degrees earned about 1.85 times as much as high school graduates in 2020, up from 1.59 times in 1995, according to U.S. Census data. Winner-take-all labor market theory argues that the salary gap between the median and the top 1 percent is not due to educational differences.

Economists use the demand and supply model to reason through the most likely causes of this shift. According to the National Center for Education Statistics, in recent decades, the supply of U.S. workers with college degrees has increased substantially. For example, 840,000 four-year bachelor's degrees were conferred on Americans in 1970. In 2018–2019, 2.0 million such degrees were conferred—an increase of over 138%. In [Figure 15.9](#), this shift in supply to the right, from S_0 to S_1 , by itself should result in a lower equilibrium wage for high-skilled labor. Thus, we can explain the increase in the price of high-skilled labor by a greater demand, like the movement from D_0 to D_1 . Evidently, combining both the increase in supply and in demand has resulted in a shift from E_0 to E_1 , and a resulting higher wage.

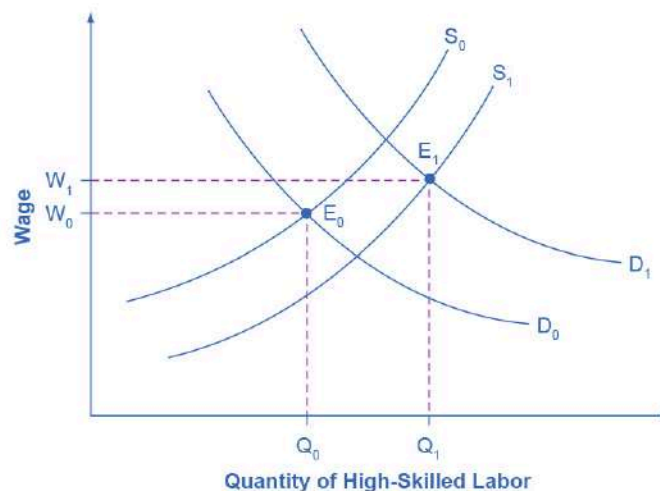


FIGURE 15.9 Why Would Wages Rise for High-Skilled Labor? The proportion of workers attending college has increased in recent decades, so the supply curve for high-skilled labor has shifted to the right, from S_0 to S_1 . If the demand for high-skilled labor had remained at D_0 , then this shift in supply would have led to lower wages for high-skilled labor. However, the wages for high-skilled labor, especially if there is a large global demand, have increased even with the shift in supply to the right. The explanation must lie in a shift to the right in demand for high-skilled labor, from D_0 to D_1 . The figure shows how a combination of the shift in supply, from S_0 to S_1 , and the shift in demand, from D_0 to D_1 , led to both an increase in the quantity of high-skilled labor hired and also to a rise in the wage for such labor, from W_0 to W_1 .

What factors would cause the demand for high-skilled labor to rise? The most plausible explanation is that while the explosion in new information and communications technologies over the last several decades has helped many workers to become more productive, the benefits have been especially great for high-skilled workers like top business managers, consultants, and design professionals. The new technologies have also helped to encourage globalization, the remarkable increase in international trade over the last few decades, by making it more possible to learn about and coordinate economic interactions all around the world. In turn, the rising impact of foreign trade in the U.S. economy has opened up greater opportunities for high-skilled workers to sell their services around the world, and lower-skilled workers have to compete with a larger supply of similarly skilled workers around the globe.

We can view the market for high-skilled labor as a race between forces of supply and demand. Additional education and on-the-job training will tend to increase the high-skilled labor supply and to hold down its relative wage. Conversely, new technology and other economic trends like globalization tend to increase the demand for high-skilled labor and push up its relative wage. We can view the greater inequality of wages as a sign that demand for skilled labor is increasing faster than supply. Alternatively, if the supply of lower skilled workers exceeds the demand, then average wages in the lower quintiles of the income distribution will decrease. The combination of forces in the high-skilled and low-skilled labor markets leads to increased income disparity.

15.5 Government Policies to Reduce Income Inequality

LEARNING OBJECTIVES

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

- Explain the arguments for and against government intervention in a market economy
- Identify beneficial ways to reduce the economic inequality in a society
- Show the tradeoff between incentives and income equality

No society should expect or desire complete equality of income at a given point in time, for a number of reasons. First, most workers receive relatively low earnings in their first few jobs, higher earnings as they reach middle age, and then lower earnings after retirement. Thus, a society with people of varying ages will have a certain amount of income inequality. Second, people's preferences and desires differ. Some are willing to work long hours to have income for large houses, fast cars and computers, luxury vacations, and the ability to support children and grandchildren.

These factors all imply that a snapshot of inequality in a given year does not provide an accurate picture of how people's incomes rise and fall over time. Even if we expect some degree of economic inequality at any point in time, how much inequality should there be? There is also the difference between income and wealth, as the following Clear It Up feature explains.



CLEAR IT UP

How do you measure wealth versus income inequality?

Income is a flow of money received, often measured on a monthly or an annual basis. **Wealth** is the sum of the value of all assets, including money in bank accounts, financial investments, a pension fund, and the value of a home. In calculating wealth, one must subtract all debts, such as debt owed on a home mortgage and on credit cards. A retired person, for example, may have relatively little income in a given year, other than a pension or Social Security. However, if that person has saved and invested over time, the person's accumulated wealth can be quite substantial.

In the United States, the wealth distribution is more unequal than the income distribution, because differences in income can accumulate over time to make even larger differences in wealth. However, we can measure the degree of inequality in the wealth distribution with the same tools we use to measure the inequality in the income distribution, like quintile measurements. Once every three years the Federal Reserve Bank publishes the *Survey of Consumer Finance* which reports a collection of data on wealth.

Even if they cannot answer the question of how much inequality is too much, economists can still play an important role in spelling out policy options and tradeoffs. If a society decides to reduce the level of economic inequality, it has three main sets of tools: redistribution from those with high incomes to those with low incomes; trying to assure that a ladder of opportunity is widely available; and a tax on inheritance.

Redistribution

Redistribution means taking income from those with higher incomes and providing income to those with

lower incomes. Earlier in this chapter, we considered some of the key government policies that provide support for people experiencing poverty: the welfare program TANF, the earned income tax credit, SNAP, and Medicaid. If a reduction in inequality is desired, these programs could receive additional funding.

The federal income tax, which is a **progressive tax system** designed in such a way that the rich pay a higher percent in income taxes than the poor, funds the programs. Data from household income tax returns in 2018 shows that the top 1% of households had an average income of \$1,679,000 per year in pre-tax income and paid an average federal tax rate of 25.4%. The **effective income tax**, which is total taxes paid divided by total income (all sources of income such as wages, profits, interest, rental income, and government transfers such as veterans' benefits), was much lower. The effective tax paid by that top 1% of householders paid was 20.4%, while the bottom two quintiles actually paid negative effective income taxes, because of provisions like the earned income tax credit. News stories occasionally report on a high-income person who has managed to pay very little in taxes, but while such individual cases exist, according to the Congressional Budget Office, the typical pattern is that people with higher incomes pay a higher average share of their income in federal income taxes.

Of course, the fact that some degree of redistribution occurs now through the federal income tax and government antipoverty programs does not settle the questions of how much redistribution is appropriate, and whether more redistribution should occur.

The Ladder of Opportunity

Economic inequality is perhaps most troubling when it is not the result of effort or talent, but instead is determined by the circumstances under which a child grows up. One child attends a well-run grade school and high school and heads on to college, while parents help out by supporting education and other interests, paying for college, a first car, and a first house, and offering work connections that lead to internships and jobs. Another child attends a poorly run grade school, barely makes it through a low-quality high school, does not go to college, and lacks family and peer support. These two children may be similar in their underlying talents and in the effort they put forth, but their economic outcomes are likely to be quite different.

Public policy can attempt to build a ladder of opportunities so that, even though all children will never come from identical families and attend identical schools, each child has a reasonable opportunity to attain an economic niche in society based on their interests, desires, talents, and efforts. [Table 15.8](#) shows some of those initiatives.

Children	College Level	Adults
• Improved day care	• Widespread loans and grants for those in financial need	• Opportunities for retraining and acquiring new skills
• Enrichment programs for preschoolers	• Public support for a range of institutions from two-year community colleges to large research universities	• Prohibiting discrimination in job markets and housing on the basis of race, gender, age, and disability
• Improved public schools	-	-

TABLE 15.8 Public Policy Initiatives

Children	College Level	Adults
• After school and community activities	-	-
• Internships and apprenticeships	-	-

TABLE 15.8 Public Policy Initiatives

Some have called the United States a land of opportunity. Although the general idea of a ladder of opportunity for all citizens continues to exert a powerful attraction, specifics are often quite controversial. Society can experiment with a wide variety of proposals for building a ladder of opportunity, especially for those who otherwise seem likely to start their lives in a disadvantaged position. The government needs to carry out such policy experiments in a spirit of open-mindedness, because some will succeed while others will not show positive results or will cost too much to enact on a widespread basis.

Inheritance Taxes

There is always a debate about inheritance taxes. It goes like this: Why should people who have worked hard all their lives and saved up a substantial nest egg not be able to give their money and possessions to their children and grandchildren? In particular, it would seem un-American if children were unable to inherit a family business or a family home. Alternatively, many Americans are far more comfortable with inequality resulting from high-income people who earned their money by starting innovative new companies than they are with inequality resulting from high-income people who have inherited money from rich parents.

The United States does have an **estate tax**—that is, a tax imposed on the value of an inheritance—which suggests a willingness to limit how much wealth one can pass on as an inheritance. However, in 2022 the estate tax applied only to those leaving inheritances of more than \$12.06 million and thus applies to only a tiny percentage of those with high levels of wealth.

The Tradeoff between Incentives and Income Equality

Government policies to reduce poverty or to encourage economic equality, if carried to extremes, can injure incentives for economic output. The poverty trap, for example, defines a situation where guaranteeing a certain level of income can eliminate or reduce the incentive to work. An extremely high degree of redistribution, with very high taxes on the rich, would be likely to discourage work and entrepreneurship. Thus, it is common to draw the tradeoff between economic output and equality, as [Figure 15.10](#) (a) shows. In this formulation, if society wishes a high level of economic output, like point A, it must also accept a high degree of inequality. Conversely, if society wants a high level of equality, like point B, it must accept a lower level of economic output because of reduced incentives for production.

This view of the tradeoff between economic output and equality may be too pessimistic, and [Figure 15.10](#) (b) presents an alternate vision. Here, the tradeoff between economic output and equality first slopes up, in the vicinity of choice C, suggesting that certain programs might increase both output and economic equality. For example, the policy of providing free public education has an element of redistribution, since the value of the public schooling received by children of low-income families is clearly higher than what low-income families pay in taxes. A well-educated population, however, is also an enormously powerful factor in providing the skilled workers of tomorrow and helping the economy to grow and expand. In this case, equality and economic growth may complement each other.

Moreover, policies to diminish inequality and soften the hardship of poverty may sustain political support for a market economy. After all, if society does not make some effort toward reducing inequality and poverty, the alternative might be that people would rebel against market forces. Citizens might seek economic security by demanding that their legislators pass laws forbidding employers from ever laying off workers or reducing wages, or laws that would impose price floors and price ceilings and shut off international trade. From this viewpoint, policies to reduce inequality may help economic output by building social support for allowing markets to operate.

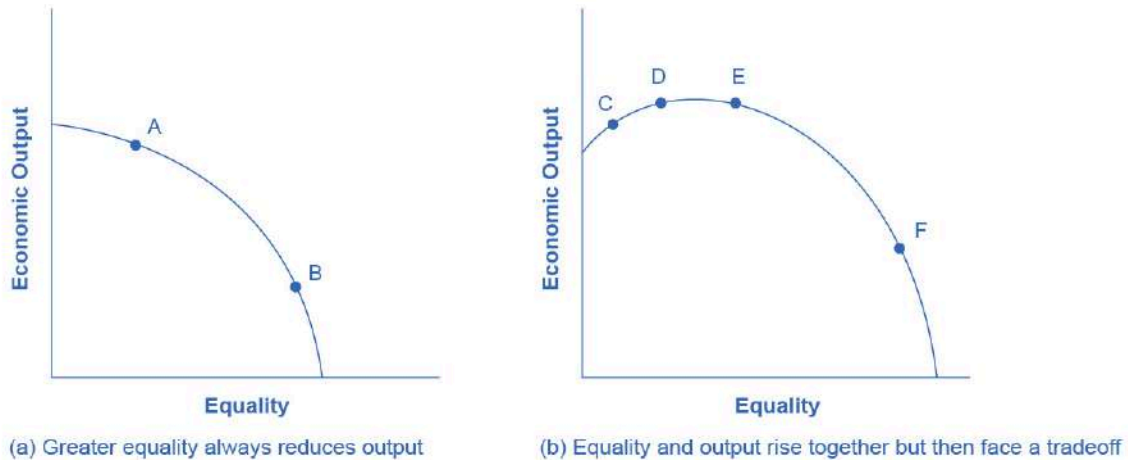


FIGURE 15.10 The Tradeoff between Incentives and Economic Equality (a) Society faces a trade-off where any attempt to move toward greater equality, like moving from choice A to B, involves a reduction in economic output. (b) Situations can arise like point C, where it is possible both to increase equality and also to increase economic output, to a choice like D. It may also be possible to increase equality with little impact on economic output, like the movement from choice D to E. However, at some point, too aggressive a push for equality will tend to reduce economic output, as in the shift from E to F.

The tradeoff in [Figure 15.10](#) (b) then flattens out in the area between points D and E, which reflects the pattern that a number of countries that provide similar levels of income to their citizens—the United States, Canada, European Union nations, Japan, and Australia—have different levels of inequality. The pattern suggests that countries in this range could choose a greater or a lesser degree of inequality without much impact on economic output. Only if these countries push for a much higher level of equality, like at point F, will they experience the diminished incentives that lead to lower levels of economic output. In this view, while a danger always exists that an agenda to reduce poverty or inequality can be poorly designed or pushed too far, it is also possible to discover and design policies that improve equality and do not injure incentives for economic output by very much—or even improve such incentives.



BRING IT HOME

Occupy Wall Street

The Occupy movement took on a life of its own over the last few months of 2011, bringing to light issues that many people faced on the lower end of the income distribution. The contents of this chapter indicate that there is a significant amount of income inequality in the United States. The question is: What should be done about it?

The 2008-2009 Great Recession caused unemployment to rise and incomes to fall. Many people attribute the recession to mismanagement of the financial system by bankers and financial managers—those in the 1% of the income distribution—but those in lower quintiles bore the greater burden of the recession through unemployment. This seemed to present the picture of inequality in a different light: the group that seemed responsible for the recession was not the group that seemed to bear the burden of the decline in output. A burden shared can bring a

society closer together. A burden pushed off onto others can polarize it.

On one level, the problem with trying to reduce income inequality comes down to whether you still believe in the American Dream. If you believe that one day you will have your American Dream—a large income, large house, happy family, or whatever else you would like to have in life—then you do not necessarily want to prevent anyone else from living out their dream. You certainly would not want to run the risk that someone would want to take part of your dream away from you. Thus, there is some reluctance to engage in a redistributive policy to reduce inequality.

However, when those for whom the likelihood of living the American Dream is very small are considered, there are sound arguments in favor of trying to create greater balance. As the text indicated, a little more income equality, gained through long-term programs like increased education and job training, can increase overall economic output. Then everyone is made better off, and the 1% will not seem like such a small group any more.

Key Terms

earned income tax credit (EITC) a method of assisting the working poor through the tax system

effective income tax percentage of total taxes paid divided by total income

estate tax a tax imposed on the value of an inheritance

income a flow of money received, often measured on a monthly or an annual basis

income inequality when one group receives a disproportionate share of total income or wealth than others

Lorenz curve a graph that compares the cumulative income actually received to a perfectly equal distribution of income; it shows the share of population on the horizontal axis and the cumulative percentage of total income received on the vertical axis

Medicaid a federal–state joint program enacted in 1965 that provides medical insurance for certain (not all) people with a low-income, including those near the poverty line as well as those below the poverty line, and focusing on low-income families with children, the low-income elderly, and people with disabilities

poverty the situation of being below a certain level of income one needs for a basic standard of living

poverty line the specific amount of income one requires for a basic standard of living

poverty rate percentage of the population living below the poverty line

poverty trap antipoverty programs set up so that government benefits decline substantially as people earn more income—as a result, working provides little financial gain

progressive tax system a tax system in which the rich pay a higher percentage of their income in taxes, rather than a higher absolute amount

quintile dividing a group into fifths, a method economists often use to look at distribution of income

redistribution taking income from those with higher incomes and providing income to those with lower incomes

safety net the group of government programs that provide assistance to people at or near the poverty line

Supplemental Nutrition Assistance Program (SNAP) a federally funded program, started in 1964, in which each month poor people receive SNAP cards they can use to buy food

wealth the sum of the value of all assets, including money in bank accounts, financial investments, a pension fund, and the value of a home

Key Concepts and Summary

15.1 Drawing the Poverty Line

Wages are influenced by Supply and demand in labor markets influence wages. This can lead to very low incomes for some people and very high incomes for others. Poverty and income inequality are not the same thing. Poverty applies to the condition of people who cannot afford the necessities of life. Income inequality refers to the disparity between those with higher and lower incomes. The poverty rate is what percentage of the population lives below the poverty line, which the amount of income that it takes to purchase the necessities of life determines. Choosing a poverty line will always be somewhat controversial.

15.2 The Poverty Trap

A poverty trap occurs when government-support payments decline as the recipients earn more income. As a result, the recipients do not end up with much more income when they work, because the loss of government support largely or completely offsets any income that one earns by working. Phasing out government benefits more slowly, as well as imposing requirements for work as a condition of receiving benefits and a time limit on benefits can reduce the harshness of the poverty trap.

15.3 The Safety Net

We call the group of government programs that address poverty the safety net. In the United States, prominent safety net programs include Temporary Assistance to Needy Families (TANF), the Supplemental Nutrition Assistance Program (SNAP), the earned income tax credit (EITC), Medicaid, and the Special Supplemental Food Program for Women, Infants, and Children (WIC).

15.4 Income Inequality: Measurement and Causes

Measuring inequality involves making comparisons across the entire distribution of income. One way of doing this is to divide the population into groups, like quintiles, and then calculate what share of income each group receives. An alternative approach is to draw Lorenz curves, which compare the cumulative income actually received to a perfectly equal distribution of income. Income inequality in the United States increased substantially from the late 1970s and early 1980s into the 2000s. The two most common explanations that economists cite are changes in household structures that have led to more two-earner couples and single-parent families, and the effect of new information and communications technology on wages.

15.5 Government Policies to Reduce Income Inequality

Policies that can affect the level of economic inequality include redistribution between rich and poor, making it easier for people to climb the ladder of opportunity; and estate taxes, which are taxes on inheritances. Pushing too aggressively for economic equality can run the risk of decreasing economic incentives. However, a moderate push for economic equality can increase economic output, both through methods like improved education and by building a base of political support for market forces.

Self-Check Questions

- Describe how each of these changes is likely to affect poverty and inequality:
 - Incomes rise for low-income and high-income workers, but rise more for the high-income earners.
 - Incomes fall for low-income and high-income workers, but fall more for high-income earners.
- Jonathon is a single father with one child. He can work as a server for \$6 per hour for up to 1,500 hours per year. He is eligible for welfare, and so if he does not earn any income, he will receive a total of \$10,000 per year. He can work and still receive government benefits, but for every \$1 of income, his welfare stipend is \$1 less. Create a table similar to [Table 15.4](#) that shows Jonathan's options. Use four columns, the first showing number of hours to work, the second showing his earnings from work, the third showing the government benefits he will receive, and the fourth column showing his total income (earnings + government support). Sketch a labor-leisure diagram of Jonathan's opportunity set with and without government support.
- Imagine that the government reworks the welfare policy that was affecting Jonathan in question 1, so that for each dollar someone like Jonathan earns at work, his government benefits diminish by only 30 cents. Reconstruct the table from question 1 to account for this change in policy. Draw Jonathan's labor-leisure opportunity sets, both for before this welfare program is enacted and after it is enacted.
- We have discovered that the welfare system discourages recipients from working because the more income they earn, the less welfare benefits they receive. How does the earned income tax credit attempt to loosen the poverty trap?
- How does the TANF attempt to loosen the poverty trap?
- A group of 10 people have the following annual incomes: \$24,000, \$18,000, \$50,000, \$100,000, \$12,000, \$36,000, \$80,000, \$10,000, \$24,000, \$16,000. Calculate the share of total income that each quintile receives from this income distribution. Do the top and bottom quintiles in this distribution have a greater or larger share of total income than the top and bottom quintiles of the U.S. income distribution?

7. [Table 15.9](#) shows the share of income going to each quintile of the income distribution for the United Kingdom in 1979 and 1991. Use this data to calculate what the points on a Lorenz curve would be, and sketch the Lorenz curve. How did inequality in the United Kingdom shift over this time period? How can you see the patterns in the quintiles in the Lorenz curves?

Share of Income	1979	1991
Top quintile	39.7%	42.9%
Fourth quintile	24.8%	22.7%
Middle quintile	17.0%	16.3%
Second quintile	11.5%	11.5%
Bottom quintile	7.0%	6.6%

TABLE 15.9 Income Distribution in the United Kingdom, 1979 and 1991

8. Using two demand and supply diagrams, one for the low-wage labor market and one for the high-wage labor market, explain how information technology can increase income inequality if it is a complement to high-income workers like salespeople and managers, but a substitute for low-income workers like file clerks and telephone receptionists.
9. Using two demand and supply diagrams, one for the low-wage labor market and one for the high-wage labor market, explain how a program that increased educational levels for a substantial number of low-skill workers could reduce income inequality.
10. Here is one hypothesis: A well-funded social safety net can increase economic equality but will reduce economic output. Explain why this might be so, and sketch a production possibility curve that shows this tradeoff.
11. Here is a second hypothesis: A well-funded social safety net may lead to less regulation of the market economy. Explain why this might be so, and sketch a production possibility curve that shows this tradeoff.
12. Which set of policies is more likely to cause a tradeoff between economic output and equality: policies of redistribution or policies aimed at the ladder of opportunity? Explain how the production possibility frontier tradeoff between economic equality and output might look in each case.
13. Why is there reluctance on the part of some in the United States to redistribute income so that greater equality can be achieved?

Review Questions

14. How is the poverty rate calculated?
15. What is the poverty line?
16. What is the difference between poverty and income inequality?
17. How does the poverty trap discourage people from working?
18. How can the effect of the poverty trap be reduced?
19. How does the U.S. government specifically support elderly people experiencing poverty?

20. What is the safety net?
21. Briefly explain the differences between TANF, the earned income tax credit, SNAP, and Medicaid.
22. Who is included in the top income quintile?
23. What is measured on the two axes of a Lorenz curve?
24. If a country had perfect income equality what would the Lorenz curve look like?
25. How has the inequality of income changed in the U.S. economy since the late 1970s?
26. What are some reasons why a certain degree of inequality of income would be expected in a market economy?
27. What are the main reasons economists give for the increase in inequality of incomes?
28. Identify some public policies that can reduce the level of economic inequality.
29. Describe how a push for economic equality might reduce incentives to work and produce output. Then describe how a push for economic inequality might not have such effects.

Critical Thinking Questions

30. What goods and services would you include in an estimate of the basic necessities for a family of four?
31. If a family of three earned \$20,000, would they be able to make ends meet given the official poverty threshold?
32. [Exercise 15.2](#) and [Exercise 15.3](#) asked you to describe the labor-leisure tradeoff for Jonathon. Since, in the first example, there is no monetary incentive for Jonathon to work, explain why he may choose to work anyway. Explain what the opportunity costs of working and not working might be for Jonathon in each example. Using your tables and graphs from [Exercise 15.2](#) and [Exercise 15.3](#), analyze how the government welfare system affects Jonathan's incentive to work.
33. Explain how you would create a government program that would give an incentive for labor to increase hours and keep labor from falling into the poverty trap.
34. Many critics of government programs to help low-income individuals argue that these programs create a poverty trap. Explain how programs such as TANF, EITC, SNAP, and Medicaid will affect low-income individuals and whether or not you think these programs will benefit families and children.
35. Think about the business cycle: during a recession, unemployment increases; it decreases in an expansionary phase. Explain what happens to TANF, SNAP, and Medicaid programs at each phase of the business cycle (recession, trough, expansion, and peak).
36. Explain how a country may experience greater equality in the distribution of income, yet still experience high rates of poverty. *Hint:* Look at the [Clear It Up](#) "How do governments measure poverty in low-income countries?" and compare to [Table 15.5](#).
37. The demand for skilled workers in the United States has been increasing. To increase the supply of skilled workers, many argue that immigration reform to allow more skilled labor into the United States is needed. Explain whether you agree or disagree.
38. Explain a situation using the supply and demand for skilled labor in which the increased number of college graduates leads to depressed wages. Given the rising cost of going to college, explain why a college education will or will not increase income inequality.
39. What do you think is more important to focus on when considering inequality: income inequality or wealth inequality?

40. To reduce income inequality, should the marginal tax rates on the top 1% be increased?
41. Redistribution of income occurs through the federal income tax and government antipoverty programs. Explain whether or not this level of redistribution is appropriate and whether more redistribution should occur.
42. How does a society or a country make the decision about the tradeoff between equality and economic output? *Hint:* Think about the political system.
43. Explain what the long- and short-term consequences are of not promoting equality or working to reduce poverty.

Problems

44. In country A, the population is 300 million and 50 million people are living below the poverty line. What is the poverty rate?
45. In country B, the population is 900 million and 100 million people are living below the poverty line. What is the poverty rate?
46. Susan is a single mother with three children. She can earn \$8 per hour and works up to 1,800 hours per year. However, if she does not earn any income at all, she will receive government benefits totaling \$16,000 per year. For every \$1 of income she earns, her level of government support will be reduced by \$1. Create a table, patterned after [this one](#). The first column should show Susan's choices of how many hours to work per year, up to 1,800 hours. The second column should show her earnings from work. The third column should show her level of government support, given her earnings. The final column should show her total income, combining earnings and government support. Based on the table you created, what are the likely impacts of this kind of assistance program on Susan's incentive to work? Are there additional opportunity costs that may reduce her incentive to work?
47. A group of 10 people have the following annual incomes: \$55,000, \$30,000, \$15,000, \$20,000, \$35,000, \$80,000, \$40,000, \$45,000, \$30,000, \$50,000. Calculate the share of total income each quintile of this income distribution received. Do the top and bottom quintiles in this distribution have a greater or larger share of total income than the top and bottom quintiles of the U.S. income distribution for 2005?



FIGURE 16.1 Former President Obama’s Health Care Reform The Patient Protection and Affordable Care Act (PPACA), more commonly known as Obamacare, relates strongly to the topic of this chapter. While originally a controversial topic, it has gained majority approval at 55% as of March 2022. (Credit: “Obama at Healthcare rally at UMD” by Daniel Borman/Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- The Problem of Imperfect Information and Asymmetric Information
- Insurance and Imperfect Information

Introduction to Information, Risk, and Insurance



BRING IT HOME

What’s the Big Deal with Obamacare?

In August 2009, many members of the U.S. Congress used their summer recess to return to their home districts and hold town hall-style meetings to discuss President Obama’s proposed changes to the U.S. healthcare system. This was officially known as the Patient Protection and Affordable Care Act (PPACA) or as the Affordable Care Act (ACA), but was more popularly known as Obamacare. The bill’s opponents’ claims ranged from the charge that the changes were unconstitutional and would add \$750 billion to the deficit, to extreme claims about the inclusion of things like the implantation of microchips and so-called “death panels” that decide which critically-ill patients receive care and which do not.

Why did people react so strongly? After all, the intent of the law is to make healthcare insurance more affordable, to allow more people to obtain insurance, and to reduce the costs of healthcare. For each year from 2000 to 2011, these costs grew at least double the rate of inflation. In 2014, healthcare spending accounted for around 24% of all federal government spending. In the United States, we spend more for our healthcare than any other high-income nation, yet our health outcomes are worse than comparable high-income countries. In 2015, over 32 million people in the United States, about 12.8% of the non-elderly adult population, were without insurance. Even today, however, more than a decade after the Act was signed into law and after the Supreme Court mostly upheld it, a 2022 Kaiser Foundation poll found that 42% of likely voters viewed it unfavorably. Why is this?

The debate over the ACA and healthcare reform could take an entire textbook, but what this chapter will do is introduce the basics of insurance and the problems insurance companies face. It is these problems, and how insurance companies respond to them that, in part, explain the divided opinion about the ACA.

Every purchase is based on a belief about the satisfaction that the good or service will provide. In turn, these beliefs are based on the information that the buyer has available. For many products, the information available to the buyer or the seller is imperfect or unclear, which can either make buyers regret past purchases or avoid making future ones.

This chapter discusses how imperfect and asymmetric information affect markets. The first module of the chapter discusses how asymmetric information affects markets for goods, labor, and financial capital. When buyers have less information about the quality of the good (for example, a gemstone) than sellers do, sellers may be tempted to mislead buyers. If a buyer cannot have at least some confidence in the quality of what they are purchasing, then they will be reluctant or unwilling to purchase the products. Thus, we require mechanisms to bridge this information gap, so buyers and sellers can engage in a transaction.

The second module of the chapter discusses insurance markets, which also face similar problems of imperfect information. For example, a car insurance company would prefer to sell insurance only to those who are unlikely to have auto accidents—but it is hard for the firm to identify those perfectly safe drivers. Conversely, car insurance buyers would like to persuade the auto insurance company that they are safe drivers and should pay only a low price for coverage. If insurance markets cannot find ways to grapple with these problems of imperfect information, then even people who have low or average risks of making claims may not be able to purchase insurance. The chapter on financial markets (markets for stocks and bonds) will show that the problems of imperfect information can be especially poignant. We cannot eliminate imperfect information, but we can often manage it.

16.1 The Problem of Imperfect Information and Asymmetric Information

LEARNING OBJECTIVES

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

- Analyze the impact of both imperfect information and asymmetric information
- Evaluate the role of advertisements in creating imperfect information
- Identify ways to reduce the risk of imperfect information
- Explain how imperfect information can affect price, quantity, and quality

Consider a purchase that many people make at important times in their lives: buying expensive jewelry. In May 1994, celebrity psychologist Doree Lynn bought an expensive ring from a jeweler in Washington, D.C., which included an emerald that cost \$14,500. Several years later, the emerald fractured. Lynn took it to another jeweler who found that cracks in the emerald had been filled with an epoxy resin. Lynn sued the original jeweler in 1997 for selling her a treated emerald without telling her, and won. The case publicized a number of little-known facts about precious stones. Most emeralds have internal flaws, and so they are soaked in clear oil or an epoxy resin to hide the flaws and make the color more deep and clear. Clear oil can leak out over time, and epoxy resin can discolor with age or heat. However, using clear oil or epoxy to “fill” emeralds is completely

legal, as long as it is disclosed.

After Doree Lynn's lawsuit, the NBC news show "Dateline" bought emeralds at four prominent jewelry stores in New York City in 1997. All the sales clerks at these stores, unaware that they were being recorded on a hidden camera, said the stones were untreated. When the emeralds were tested at a laboratory, however, technicians discovered they had all been treated with oil or epoxy. Emeralds are not the only gemstones that are treated. Diamonds, topaz, and tourmaline are also often irradiated to enhance colors. The general rule is that all treatments to gemstones should be revealed, but often sellers do not disclose this. As such, many buyers face a situation of **asymmetric information**, where two parties involved in an economic transaction have an unequal amount of information (one party knows much more than the other).

Many economic transactions occur in a situation of **imperfect information**, where either the buyer, the seller, or both, are less than 100% certain about the qualities of what they are buying and selling. Also, one may characterize the transaction as asymmetric information, in which one party has more information than the other regarding the economic transaction. Let's begin with some examples of how imperfect information complicates transactions in goods, labor, and financial capital markets. The presence of imperfect information can easily cause a decline in prices or quantities of products sold. However, buyers and sellers also have incentives to create mechanisms that will allow them to make mutually beneficial transactions even in the face of imperfect information.

If you are unclear about the difference between asymmetric information and imperfect information, read the following Clear It Up feature.



CLEAR IT UP

What is the difference between imperfect and asymmetric information?

For a market to reach equilibrium sellers and buyers must have full information about the product's price and quality. If there is limited information, then buyers and sellers may not be able to transact or will possibly make poor decisions.

Imperfect information refers to the situation where buyers and/or sellers do not have all of the necessary information to make an informed decision about the product's price or quality. The term imperfect information simply means that the buyers and/or sellers do not have all the information necessary to make an informed decision. Asymmetric information is the condition where one party, either the buyer or the seller, has more information about the product's quality or price than the other party. In either case (imperfect or asymmetric information) buyers or sellers need remedies to make more informed decisions.

"Lemons" and Other Examples of Imperfect Information

Consider Marvin, who is trying to decide whether to buy a used car. Let's assume that Marvin is truly clueless about what happens inside a car's engine. He is willing to do some background research, like reading *Consumer Reports* or checking websites that offer information about used car makes and models and what they should cost. He might pay a mechanic to inspect the car. Even after devoting some money and time collecting information, however, Marvin still cannot be absolutely sure that he is buying a high-quality used car. He knows that he might buy the car, drive it home, and use it for a few weeks before discovering that car is a "lemon," which is slang for a defective product (especially a car).

Imagine that Marvin shops for a used car and finds two that look very similar in terms of mileage, exterior appearances, and age. One car costs \$4,000, while the other car costs \$4,600. Which car should Marvin buy?

If Marvin were choosing in a world of perfect information, the answer would be simple: he should buy the cheaper car. However, Marvin is operating in a world of imperfect information, where the sellers likely know more about the car's problems than he does, and have an incentive to hide the information. After all, the more

problems the sellers disclose, the lower the car's selling price.

What should Marvin do? First, he needs to understand that even with imperfect information, prices still reflect information. Typically, used cars are more expensive on some dealer lots because the dealers have a trustworthy reputation to uphold. Those dealers try to fix problems that may not be obvious to their customers, in order to create good word of mouth about their vehicles' long term reliability. The short term benefits of selling their customers a "lemon" could cause a quick collapse in the dealer's reputation and a loss of long term profits. On other lots that are less well-established, one can find cheaper used cars, but the buyer takes on more risk when a dealer's reputation has little at stake. The cheapest cars of all often appear on Craigslist, where the individual seller has no reputation to defend. In sum, cheaper prices do carry more risk, so Marvin should balance his appetite for risk versus the potential headaches of many more unanticipated trips to the repair shop.

Similar problems with imperfect information arise in labor and financial capital markets. Consider Greta, who is applying for a job. Her potential employer, like the used car buyer, is concerned about ending up with a "lemon"—in this case a poor quality employee. The employer will collect information about Greta's academic and work history. In the end, however, a degree of uncertainty will inevitably remain regarding Greta's abilities, which are hard to demonstrate without actually observing her on the job. How can a potential employer screen for certain attributes, such as motivation, timeliness, and ability to get along with others? Employers often look to trade schools and colleges to pre-screen candidates. Employers may not even interview a candidate unless he has a degree and, sometimes, a degree from a particular school. Employers may also view awards, a high grade point average, and other accolades as a signal of hard work, perseverance, and ability. Employers may also seek references for insights into key attributes such as energy level and work ethic.

How Imperfect Information Can Affect Equilibrium Price and Quantity

The presence of imperfect information can discourage both buyers and sellers from participating in the market. Buyers may become reluctant to participate because they cannot determine the product's quality. Sellers of high-quality or medium-quality goods may be reluctant to participate, because it is difficult to demonstrate the quality of their goods to buyers—and since buyers cannot determine which goods have higher quality, they are likely to be unwilling to pay a higher price for such goods.

Economists sometimes refer to a market with few buyers and few sellers as a thin market. By contrast, they call a market with many buyers and sellers a thick market. When imperfect information is severe and buyers and sellers are discouraged from participating, markets may become extremely thin as a relatively small number of buyer and sellers attempt to communicate enough information that they can agree on a price.

When Price Mixes with Imperfect Information about Quality

A buyer confronted with imperfect information will often believe that the price reveals something about the product's quality. For example, a buyer may assume that a gemstone or a used car that costs more must be of higher quality, even though the buyer is not an expert on gemstones. Think of the expensive restaurant where the food must be good because it is so expensive or the shop where the clothes must be stylish because they cost so much, or the gallery where the art must be great, because the price tags are high. If you are hiring a lawyer, you might assume that a lawyer who charges \$400 per hour must be better than a lawyer who charges \$150 per hour. In these cases, price can act as a signal of quality.

When buyers use the market price to draw inferences about the products' quality, then markets may have trouble reaching an equilibrium price and quantity. Imagine a situation where a used car dealer has a lot full of used cars that do not seem to be selling, and so the dealer decides to cut the car prices to sell a greater quantity. In a market with imperfect information, many buyers may assume that the lower price implies low-quality cars. As a result, the lower price may not attract more customers. Conversely, a dealer who raises prices may find that customers assume that the higher price means that cars are of higher quality. As a result of raising prices, the dealer might sell more cars. (Whether or not consumers always behave rationally, as an

economist would see it, is the subject of the following Clear It Up feature.)

The idea that higher prices might cause a greater quantity demanded and that lower prices might cause a lower quantity demanded runs exactly counter to the basic model of demand and supply (as we outlined in the [Demand and Supply](#) chapter). These contrary effects, however, will reach natural limits. At some point, if the price is high enough, the quantity demanded will decline. Conversely, when the price declines far enough, buyers will increasingly find value even if the quality is lower. In addition, information eventually becomes more widely known. An overpriced restaurant that charges more than the quality of its food is worth to many buyers will not last forever.



CLEAR IT UP

Is consumer behavior rational?

There is much human behavior that mainstream economists have tended to call “irrational” since it is consistently at odds with economists’ utility maximizing models. The typical response is for economists to brush these behaviors aside and call them “anomalies” or unexplained quirks.

“If only you knew more economics, you would not be so irrational,” is what many mainstream economists seem to be saying. A group known as behavioral economists has challenged this notion, because so much of this so-called “quirky” behavior is extremely common among us. For example, a conventional economist would say that if you lost a \$10 bill today, and also received 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 conducted research that shows many people will feel some negative emotion—anger or frustration—after those two things happen. We tend to focus more on the loss than the gain. Economists Daniel Kahneman and Amos Tversky in a famous 1979 *Econometrica* paper called this “loss aversion”, where a \$1 loss pains us 2.25 times more than a \$1 gain helps us. This has implications for investing, as people tend to “overplay” the stock market by reacting more to losses than to gains.

Behavioral economics also tries to explain why people make seemingly irrational decisions in the presence of different situations, or how they “frame” the decision. We outline a popular example here: Imagine you have the opportunity to buy an alarm clock for \$20 in Store A. Across the street, you learn, is the exact same clock at Store B for \$10. You might say it is worth your time—a five-minute walk—to save \$10. Now, take a different example: You are in Store A buying a \$300 phone. Five minutes away, at Store B, the same phone is \$290. You again save \$10 by taking a five-minute walk. Do you do it?

Surprisingly, it is likely that you would not. Mainstream economists would say “\$10 is \$10” and that it would be irrational to make a five minute walk for \$10 in one case and not the other. However, behavioral economists have pointed out that most of us evaluate outcomes relative to a reference point—here the cost of the product—and think of gains and losses as percentages rather than using actual savings.

Which view is right? Both have their advantages, but behavioral economists have at least shed a light on trying to describe and explain systematic behavior which some previously had 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.

Mechanisms to Reduce the Risk of Imperfect Information

If you were selling a good like emeralds or used cars where imperfect information is likely to be a problem, how could you reassure possible buyers? If you were buying a good where imperfect information is a problem, what would it take to reassure you? Buyers and sellers in the goods market rely on reputation as well as guarantees, warranties, and service contracts to assure product quality. The labor market uses occupational licenses and certifications to assure competency, while the financial capital market uses cosigners and

collateral as insurance against unforeseen, detrimental events.

In the goods market, the seller might offer a **money-back guarantee**, an agreement that functions as a promise of quality. This strategy may be especially important for a company that sells goods through mail-order catalogs or over the web, whose customers cannot see the actual products, because it encourages people to buy something even if they are not certain they want to keep it.

L.L. Bean started using money-back-guarantees in 1911, when the founder stitched waterproof shoe rubbers together with leather shoe tops, and sold them as hunting shoes. He guaranteed satisfaction. However, the stitching came apart and, out of the first batch of 100 pairs that were sold, customers returned 90 pairs. L.L. Bean took out a bank loan, repaired all of the shoes, and replaced them. The L.L. Bean reputation for customer satisfaction began to spread. Many firms today offer money-back-guarantees for a few weeks or months, but L.L. Bean offers a complete money-back guarantee. Customers can always return anything they have bought from L.L. Bean, no matter how many years later or what condition the product is in, for a full money-back guarantee.

L.L. Bean has very few stores. Instead, most of its sales are made by mail, telephone, or, now, through their website. For this kind of firm, imperfect information may be an especially difficult problem, because customers cannot see and touch what they are buying. A combination of a money-back guarantee and a reputation for quality can help for a mail-order firm to flourish.

Sellers may offer a **warranty**, which is a promise to fix or replace the good, at least for a certain time period. The seller may also offer a buyer a chance to buy a **service contract**, where the buyer pays an extra amount and the seller agrees to fix anything that goes wrong for a set time period. Service contracts are often an option for buyers of large purchases such as cars, appliances and even houses.

Guarantees, warranties, and service contracts are examples of explicit reassurance that sellers provide. In many cases, firms also offer unstated guarantees. For example, some movie theaters might refund the ticket cost to a customer who walks out complaining about the show. Likewise, while restaurants do not generally advertise a money-back guarantee or exchange policies, many restaurants allow customers to exchange one dish for another or reduce the price of the bill if the customer is not satisfied.

The rationale for these policies is that firms want repeat customers, who in turn will recommend the business to others. As such, establishing a good reputation is of paramount importance. When buyers know that a firm is concerned about its reputation, they are less likely to worry about receiving a poor-quality product. For example, a well-established grocery store with a good reputation can often charge a higher price than a temporary stand at a local farmer's market, where the buyer may never see the seller again.

Sellers of labor provide information through resumes, recommendations, school transcripts, and examples of their work. The labor market also uses **occupational licenses** to establish quality in the labor market. Occupational licenses, which government agencies typically issue, show that a worker has completed a certain type of education or passed a certain test. Some of the professionals who must hold a license are doctors, teachers, nurses, engineers, accountants, and lawyers. In addition, most states require a license to work as a barber, an embalmer, a dietitian, a massage therapist, a hearing aid dealer, a counselor, an insurance agent, and a real estate broker. Some other jobs require a license in only one state. Minnesota requires a state license to be a field archaeologist. North Dakota has a state license for bait retailers. In Louisiana, one needs a state license to be a "stress analyst" and California requires a state license to be a furniture upholsterer. According to a 2013 study from the University of Chicago, about 29% of U.S. workers have jobs that require occupational licenses.

Occupational licenses have their downside as well, as they represent a barrier to entry to certain industries. This makes it more difficult for new entrants to compete with incumbents, which can lead to higher prices and less consumer choice. In occupations that require licenses, the government has decided that the additional information provided by licenses outweighs the negative effect on competition.



CLEAR IT UP

Are advertisers allowed to benefit from imperfect information?

Many advertisements seem full of imperfect information—at least by what they imply. Driving a certain car, drinking a particular soda, or wearing a certain shoe are all unlikely to bring fashionable friends and fun automatically, if at all. The government rules on advertising, enforced by the Federal Trade Commission (FTC), allow advertising to contain a certain amount of exaggeration about the general delight of using a product. They, however, also demand that if one presents a claim as a fact, it must be true.

Legally, deceptive advertising dates back to the 1950s when Colgate-Palmolive created a television advertisement that seemed to show Rapid Shave shaving cream being spread on sandpaper and then the sand was shaved off the sandpaper. What the television advertisement actually showed was sand sprinkled on Plexiglas—without glue—and then scraped aside by the razor.

In the 1960s, in magazine advertisements for Campbell's vegetable soup, the company was having problems getting an appetizing soup picture, because the vegetables kept sinking. To remedy this, they filled a bowl with marbles and poured the soup over the top, so that the bowl appeared to be crammed with vegetables.

In the late 1980s, the Volvo Company filmed a television advertisement that showed a monster truck driving over cars, crunching their roofs—all except for the Volvo, which did not crush. However, the FTC found in 1991 that the Volvo's roof from the filming had been reinforced with an extra steel framework, while they cut the roof supports on the other car brands.

The Wonder Bread Company ran television advertisements featuring “Professor Wonder,” who said that because Wonder Bread contained extra calcium, it would help children's minds work better and improve their memory. The FTC objected, and in 2002 the company agreed to stop running the advertisements.

As we can see in each of these cases, the Federal Trade Commission (FTC) often checks factual claims about the product's performance, at least to some extent. Language and images that are exaggerated or ambiguous, but not actually false, are allowed in advertising. Untrue “facts” are not permitted. In any case, an old Latin saying applies when watching advertisements: *Caveat emptor*—that is, “let the buyer beware.”

On the buyer's side of the labor market, a standard precaution against hiring a “lemon” of an employee is to specify that the first few months of employment are officially a trial or probationary period, and that the employer can dismiss the worker for any reason or no reason during that time. Sometimes workers also receive lower pay during this trial period.

In the financial capital market, before a bank makes a loan, it requires a prospective borrower to fill out forms regarding incomes sources. In addition, the bank conducts a credit check on the individual's past borrowing. Another approach is to require a **cosigner** on a loan; that is, another person or firm who legally pledges to repay some or all of the money if the original borrower does not do so. Another approach is to require **collateral**, often property or equipment that the bank would have a right to seize and sell if borrower does not repay the loan.

Buyers of goods and services cannot possibly become experts in evaluating the quality of gemstones, used cars, lawyers, and everything else they buy. Employers and lenders cannot be perfectly omniscient about whether possible workers will turn out well or potential borrowers will repay loans on time. However, the mechanisms that we mentioned above can reduce the risks associated with imperfect information so that the buyer and seller are willing to proceed.

16.2 Insurance and Imperfect Information

LEARNING OBJECTIVES

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

- Explain how insurance works
- Identify and evaluate various forms of government and social insurance
- Discuss the problems caused by moral hazard and adverse selection
- Analyze the impact of government regulation of insurance

Insurance is a method that households and firms use to prevent any single event from having a significant detrimental financial effect. Generally, households or firms with insurance make regular payments, called **premiums**. The insurance company prices these premiums based on the probability of certain events occurring among a pool of people. Members of the group who then suffer a specified bad experience receive payments from this pool of money.

Many people have several kinds of insurance: health insurance that pays when they receive medical care; car insurance that pays if their car is in an automobile accident; house or renter's insurance that pays for stolen possessions or items damaged by fire; and life insurance, which pays for the family if the insured individual dies. [Table 16.1](#) lists a set of insurance markets.

Type of Insurance	Who Pays for It?	It Pays Out When . . .
Health insurance	Employers and individuals	Medical expenses are incurred
Life insurance	Employers and individuals	Policyholder dies
Automobile insurance	Individuals	Car is damaged, stolen, or causes damage to others
Property and homeowner's insurance	Homeowners and renters	Dwelling is damaged or burglarized
Liability insurance	Firms and individuals	An injury occurs for which you are partly responsible
Malpractice insurance	Doctors, lawyers, and other professionals	A poor quality of service is provided that causes harm to others

TABLE 16.1 Some Insurance Markets

All insurance involves imperfect information in both an obvious way and in a deeper way. At an obvious level, we cannot predict future events with certainty. For example, we cannot know with certainty who will have a car accident, become ill, die, or have his home robbed in the next year. Imperfect information also applies to estimating the risk that something will happen to any individual. It is difficult for an insurance company to estimate the risk that, say, a particular 20-year-old male driver from New York City will have an accident, because even within that group, some drivers will drive more safely than others. Thus, adverse events occur out of a combination of people's characteristics and choices that make the risks higher or lower and then the good or bad luck of what actually happens.

How Insurance Works

A simplified example of automobile insurance might work this way. Suppose we divide a group of 100 drivers into three groups. In a given year, 60 of those people have only a few door dings or chipped paint, which costs \$100 each. Another 30 of the drivers have medium-sized accidents that cost an average of \$1,000 in damages,

and 10 of the drivers have large accidents that cost \$15,000 in damages. For the moment, let's imagine that at the beginning of any year, there is no way of identifying the drivers who are low-risk, medium-risk, or high-risk. The total damage incurred by car accidents in this group of 100 drivers will be \$186,000, that is:

$$\begin{aligned}\text{Total damage} &= (60 \times \$100) + (30 \times \$1,000) + (10 \times \$15,000) \\ &= \$6,000 + \$30,000 + \$150,000 \\ &= \$186,000\end{aligned}$$

If each of the 100 drivers pays a \$1,860 premium each year, the insurance company will collect the \$186,000 that is needed to cover the costs of the accidents that occur.

Since insurance companies have such a large number of clients, they are able to negotiate with health care and other service providers for lower rates than the individual would be able to get, thus increasing the benefit to consumers of becoming insured and saving the insurance company itself money when it pays out claims.

Insurance companies receive income, as [Figure 16.2](#) shows, from insurance premiums and investment income. The companies derive income from investing the funds that insurance companies received in the past but did not pay out as insurance claims in prior years. The insurance company receives a rate of return from investing these funds or reserves. The companies typically invest in fairly safe, liquid (easy to convert into cash) investments, as the insurance companies need to be able to readily access these funds when a major disaster strikes.



FIGURE 16.2 An Insurance Company: What Comes In, What Goes Out Money flows into an insurance company through premiums and investments and out through the payment of claims and operating expenses.

Government and Social Insurance

Federal and state governments run a number of insurance programs. Some of the programs look much like private insurance, in the sense that the members of a group make steady payments into a fund, and those in the group who suffer an adverse experience receive payments. Other programs protect against risk, but without an explicit fund set up. Following are some examples.

- **Unemployment insurance:** Employers in every state pay a small amount for unemployment insurance, which goes into a fund to pay benefits to workers who lose their jobs and do not find new jobs, for a period of time, usually up to six months.
- **Pension insurance:** Employers that offer pensions to their retired employees are required by law to pay a small fraction of what they are setting aside for pensions to the Pension Benefit Guarantee Corporation, which pays at least some pension benefits to workers if a company goes bankrupt and cannot pay the pensions it has promised.
- **Deposit insurance:** Banks are required by law to pay a small fraction of their deposits to the Federal Deposit Insurance Corporation, which goes into a fund that pays depositors the value of their bank deposits up to \$250,000 (the amount was raised from \$100,000 to \$250,000 in 2008) if the bank should go bankrupt.
- **Workman's compensation insurance:** Employers are required by law to pay a small percentage of the salaries that they pay into funds, typically run at the state level, that pay benefits to workers who suffer an injury on the job.
- **Retirement insurance:** All workers pay a percentage of their income into Social Security and into Medicare, which then provides income and health care benefits to the elderly. Social Security and Medicare are not

literally “insurance” in the sense that those currently contributing to the fund are not eligible for benefits. They function like insurance, however, in the sense that individuals make regular payments into the programs today in exchange for benefits they will receive in the case of a later event—either becoming old or becoming sick when old. A name for such programs is “social insurance.”

The major additional costs to insurance companies, other than the payment of claims, are the costs of running a business: the administrative costs of hiring workers, administering accounts, and processing insurance claims. For most insurance companies, the insurance premiums coming in and the claims payments going out are much larger than the amounts earned by investing money or the administrative costs.

Thus, while factors like investment income earned on reserves, administrative costs, and groups with different risks complicate the overall picture, a fundamental law of insurance must hold true: The average person’s payments into insurance over time must cover 1) the average person’s claims, 2) the costs of running the company, and 3) leave room for the firm’s profits.

Risk Groups and Actuarial Fairness

Not all of those who purchase insurance face the same risks. Some people may be more likely, because of genetics or personal habits, to fall sick with certain diseases. Some people may live in an area where car theft or home robbery is more likely than in other areas. Some drivers are safer than others. A **risk group** can be defined as a group that shares roughly the same risks of an adverse event occurring.

Insurance companies often classify people into risk groups, and charge lower premiums to those with lower risks. If people are not separated into risk groups, then those with low risk must pay for those with high risks. In the simple example of how car insurance works, 60 drivers had very low damage of \$100 each, 30 drivers had medium-sized accidents that cost \$1,000 each, and 10 of the drivers had large accidents that cost \$15,000. If all 100 of these drivers pay the same \$1,860, then those with low damages are in effect paying for those with high damages.

If it is possible to classify drivers according to risk group, then the insurance company can charge each group according to its expected losses. For example, the insurance company might charge the 60 drivers who seem safest of all \$100 apiece, which is the average value of the damages they cause. Then the intermediate group could pay \$1,000 apiece and the high-cost group \$15,000 each. When the level of insurance premiums that someone pays is equal to the amount that an average person in that risk group would collect in insurance payments, the level of insurance is said to be “actuarially fair.”

Classifying people into risk groups can be controversial. For example, if someone had a major automobile accident last year, should the insurance company classify that person as a high-risk driver who is likely to have similar accidents in the future, or as a low-risk driver who was just extremely unlucky? The driver is likely to claim to be low-risk, and thus someone who should be in a risk group with those who pay low insurance premiums in the future. The insurance company is likely to believe that, on average, having a major accident is a signal of being a high-risk driver, and thus try to charge this driver higher insurance premiums. The next two sections discuss the two major problems of imperfect information in insurance markets—called moral hazard and adverse selection. Both problems arise from attempts to categorize those purchasing insurance into risk groups.

The Moral Hazard Problem

Moral hazard refers to the case when people engage in riskier behavior with insurance than they would if they did not have insurance. For example, if you have health insurance that covers the cost of visiting the doctor, you may be less likely to take precautions against catching an illness that might require a doctor’s visit. If you have car insurance, you will worry less about driving or parking your car in ways that make it more likely to get dented. In another example, a business without insurance might install absolute top-level security and fire sprinkler systems to guard against theft and fire. If it is insured, that same business might only install a

minimum level of security and fire sprinkler systems.

We cannot eliminate moral hazard, but insurance companies have some ways of reducing its effect. Investigations to prevent insurance fraud are one way of reducing the extreme cases of moral hazard. Insurance companies can also monitor certain kinds of behavior. To return to the example from above, they might offer a business a lower rate on property insurance if the business installs a top-level security and fire sprinkler system and has those systems inspected once a year.

Another method to reduce moral hazard is to require the injured party to pay a share of the costs. For example, insurance policies often have **deductibles**, which is an amount that the insurance policyholder must pay out of their own pocket before the insurance coverage starts paying. For example, auto insurance might pay for all losses greater than \$500. Health insurance policies often have a **copayment**, in which the policyholder must pay a small amount. For example, a person might have to pay \$20 for each doctor visit, and the insurance company would cover the rest. Another method of cost sharing is **coinsurance**, which means that the insurance company covers a certain percentage of the cost. For example, insurance might pay for 80% of the costs of repairing a home after a fire, but the homeowner would pay the other 20%.

All of these forms of cost sharing discourage moral hazard, because people know that they will have to pay something out of their own pocket when they make an insurance claim. The effect can be powerful. One prominent study found that when people face moderate deductibles and copayments for their health insurance, they consume about one-third less in medical care than people who have complete insurance and do not pay anything out of pocket, presumably because deductibles and copayments reduce the level of moral hazard. However, those who consumed less health care did not seem to have any difference in health status.

A final way of reducing moral hazard, which is especially applicable to health care, is to focus on healthcare incentives of providers rather than consumers. Traditionally, most health care in the United States has been provided on a **fee-for-service** basis, which means that medical care providers are paid for the services they provide and are paid more if they provide additional services. However, in the last decade or so, the structure of healthcare provision has shifted to an emphasis on health maintenance organizations (HMOs). A **health maintenance organization (HMO)** provides healthcare that receives a fixed amount per person enrolled in the plan—regardless of how many services are provided. In this case, a patient with insurance has an incentive to demand more care, but the healthcare provider, which is receiving only a fixed payment, has an incentive to reduce the moral hazard problem by limiting the quantity of care provided—as long as it will not lead to worse health problems and higher costs later. Today, many doctors are paid with some combination of managed care and fee-for-service; that is, a flat amount per patient, but with additional payments for the treatment of certain health conditions.

Imperfect information is the cause of the moral hazard problem. If an insurance company had perfect information on risk, it could simply raise its premiums every time an insured party engages in riskier behavior. However, an insurance company cannot monitor all the risks that people take all the time and so, even with various checks and cost sharing, moral hazard will remain a problem.

LINK IT UP

Visit this [website \(http://openstax.org/l/healtheconomics\)](http://openstax.org/l/healtheconomics) to read about the relationship between health care and behavioral economics.

The Adverse Selection Problem

Adverse selection refers to the problem in which insurance buyers have more information about whether they are high-risk or low-risk than the insurance company does. This creates an asymmetric information problem for the insurance company because buyers who are high-risk tend to want to buy more insurance, without letting the insurance company know about their higher risk. For example, someone purchasing health

insurance or life insurance probably knows more about their family's health history than an insurer can reasonably find out even with a costly investigation. Someone purchasing car insurance may know that they are a high-risk driver who has not yet had a major accident—but it is hard for the insurance company to collect information about how people actually drive.

To understand how adverse selection can strangle an insurance market, recall the situation of 100 drivers who are buying automobile insurance, where 60 drivers had very low damages of \$100 each, 30 drivers had medium-sized accidents that cost \$1,000 each, and 10 of the drivers had large accidents that cost \$15,000. That would equal \$186,000 in total payouts by the insurance company. Imagine that, while the insurance company knows the overall size of the losses, it cannot identify the high-risk, medium-risk, and low-risk drivers. However, the drivers themselves know their risk groups. Since there is asymmetric information between the insurance company and the drivers, the insurance company would likely set the price of insurance at \$1,860 per year, to cover the average loss (not including the cost of overhead and profit). The result is that those with low risks of only \$100 will likely decide not to buy insurance; after all, it makes no sense for them to pay \$1,860 per year when they are likely only to experience losses of \$100. Those with medium risks of a \$1,000 accident will not buy insurance either. Therefore, the insurance company ends up only selling insurance for \$1,860 to high-risk drivers who will average \$15,000 in claims apiece, and as a consequence, the insurance company ends up losing considerable money. If the insurance company tries to raise its premiums to cover the losses of those with high risks, then those with low or medium risks will be even more discouraged from buying insurance.

Rather than face such a situation of adverse selection, the insurance company may decide not to sell insurance in this market at all. If potential buyers are to receive insurance, then one of two things must happen. First, the insurance company might find some way of separating insurance buyers into risk groups with some degree of accuracy and charging them accordingly, which in practice often means that the insurance company tries not to sell insurance to those who may pose high risks. Another scenario is that those with low risks must buy insurance, even if they have to pay more than the actuarially fair amount for their risk group. The notion that people can be required to purchase insurance raises the issue of government laws and regulations that influence the insurance industry.

U.S. Health Care in an International Context

The United States is the only high-income country in the world where private firms pay and provide for most health insurance. Greater government involvement in the provision of health insurance is one possible way of addressing moral hazard and adverse selection problems.

The moral hazard problem with health insurance is that when people have insurance, they will demand higher quantities of health care. In the United States, private healthcare insurance tends to encourage an ever-greater demand for healthcare services, which healthcare providers are happy to fulfill. [Table 16.2](#) shows that on a per-person basis, U.S. healthcare spending towers above healthcare spending of other countries. Note that while healthcare expenditures in the United States are far higher than healthcare expenditures in other countries, the health outcomes in the United States, as measured by life expectancy and lower rates of childhood mortality, tend to be lower. Health outcomes, however, may not be significantly affected by healthcare expenditures. Many studies have shown that a country's health is more closely related to diet, exercise, and genetic factors than to healthcare expenditure. This fact further emphasizes that the United States is spending very large amounts on medical care with little obvious health gain.

In the U.S. health insurance market, the main way of solving this adverse selection problem is that health insurance is often sold through groups based on place of employment, or, under The Affordable Care Act, from a state government sponsored health exchange market. From an insurance company's point of view, selling insurance through an employer mixes together a group of people—some with high risks of future health problems and some with lower risks—and thus reduces the insurance firm's fear of attracting only those who have high risks. However, many small companies do not provide health insurance to their employees, and

many lower-paying jobs do not include health insurance. Even after we take into account all U.S. government programs that provide health insurance for the elderly and people experiencing poverty, approximately 31 million Americans were without health insurance in 2020. While a government-controlled system can avoid the adverse selection problem entirely by providing at least basic health insurance for all, another option is to mandate that all Americans buy health insurance from some provider by preventing providers from denying individuals based on preexisting conditions. The Patient Protection and Affordable Care Act adopted this approach, which we will discuss later on in this chapter.

Country	Health Care Spending per Person	Life Expectancy at Birth (Male)	Life Expectancy at Birth (Female)	Infant Mortality Rate (Male & Female), per 1,000
United States	\$10,948	75.5	80.2	5.7
Germany	\$6,731	79.0	83.7	3.2
France	\$5,564	79.2	85.3	3.5
Canada	\$5,370	80.0	84.2	4.4
United Kingdom	\$5,268	78.4	82.4	3.7

TABLE 16.2 A Comparison of Health Care Spending per Person, Life Expectancy at Birth, and Infant Mortality, 2020 (Source: 2020 OECD study and World Fact Book)

At its best, the largely private U.S. system of health insurance and healthcare delivery provides an extraordinarily high quality of care, along with generating a seemingly endless parade of life-saving innovations. However, the system also struggles to control its high costs and to provide basic medical care to all. Compared to the United States, other countries have lower costs, more equal access, and better mortality outcomes, but they often struggle to provide rapid access to health care and to offer the near-miracles of the most up-to-date medical care. The challenge is a healthcare system that strikes the right balance between quality, access, and cost.

Government Regulation of Insurance

The U.S. insurance industry is primarily regulated at the state level. Since 1871 there has been a National Association of Insurance Commissioners that brings together these state regulators to exchange information and strategies. The state insurance regulators typically attempt to accomplish two things: to keep the price of insurance low and to ensure that everyone has insurance. These goals, however, can conflict with each other and also become easily entangled in politics.

If insurance premiums are set at actuarially fair levels, so that people end up paying an amount that accurately reflects their risk group, certain people will end up paying considerable amounts. For example, if health insurance companies were trying to cover people who already have a chronic disease like AIDS, or who were elderly, they would charge these groups very high premiums for health insurance, because their expected health care costs are quite high. Women in the age bracket 18–44 consume, on average, about 65% more in health care spending than men. Young male drivers have more car accidents than young female drivers. Thus, actuarially fair insurance would tend to charge young men much more for car insurance than young women. Because people in high-risk groups would find themselves charged so heavily for insurance, they might choose not to buy insurance at all.

State insurance regulators have sometimes reacted by passing rules that attempt to set low premiums for insurance. Over time, however, the fundamental law of insurance must hold: the average amount individuals receive cannot exceed the average amount paid in premiums. When rules are passed to keep premiums low, insurance companies try to avoid insuring any high-risk or even medium-risk parties. If a state legislature passes strict rules requiring insurance companies to sell to everyone at low prices, the insurance companies always have the option of withdrawing from doing business in that state. For example, the insurance regulators in New Jersey are well-known for attempting to keep auto insurance premiums low, and more than 20 different insurance companies stopped doing business in the state in the late 1990s and early 2000s. Similarly, in 2009, State Farm announced that it was withdrawing from selling property insurance in Florida.

In short, government regulators cannot force companies to charge low prices and provide high levels of insurance coverage—and thus take losses—for a sustained period of time. If insurance premiums are set below the actuarially fair level for a certain group, some other group will have to make up the difference. There are two other groups who can make up the difference: taxpayers or other insurance buyers.

In some industries, the U.S. government has decided free markets will not provide insurance at an affordable price, and so the government pays for it directly. For example, private health insurance is too expensive for many people whose incomes are too low. To combat this, the U.S. government, together with the states, runs the Medicaid program, which provides health care to those with low incomes. Private health insurance also does not work well for the elderly, because their average health care costs can be very high. Thus, the U.S. government started the Medicare program, which provides health insurance to all those over age 65. Other government-funded health-care programs are aimed at military veterans, as an added benefit, and children in families with relatively low incomes.

Another common government intervention in insurance markets is to require that everyone buy certain kinds of insurance. For example, most states legally require car owners to buy auto insurance. Likewise, when a bank loans someone money to buy a home, the person is typically required to have homeowner's insurance, which protects against fire and other physical damage (like hailstorms) to the home. A legal requirement that everyone must buy insurance means that insurance companies do not need to worry that those with low risks will avoid buying insurance. Since insurance companies do not need to fear adverse selection, they can set their prices based on an average for the market, and those with lower risks will, to some extent, end up subsidizing those with higher risks. However, even when laws are passed requiring people to purchase insurance, insurance companies cannot be compelled to sell insurance to everyone who asks—at least not at low cost. Thus, insurance companies will still try to avoid selling insurance to those with high risks whenever possible.

The government cannot pass laws that make the problems of moral hazard and adverse selection disappear, but the government can make political decisions that certain groups should have insurance, even though the private market would not otherwise provide that insurance. Also, the government can impose the costs of that decision on taxpayers or on other buyers of insurance.

The Patient Protection and Affordable Care Act

In March of 2010, President Obama signed into law the Patient Protection and Affordable Care Act (PPACA). The government started to phase in this highly contentious law over time starting in October of 2013. The goal of the act is to bring the United States closer to universal coverage. Some of the key features of the plan include:

- **Individual mandate:** All individuals, who do not receive health care through their employer or through a government program (for example, Medicare), were required to have health insurance or pay a fine. The individual mandate's goal was to reduce the adverse selection problem and keep prices down by requiring all consumers—even the healthiest ones—to have health insurance. Without the need to guard against adverse selection (whereby only the riskiest consumers buy insurance) by raising prices, health insurance companies could provide more reasonable plans to their customers. At the beginning of 2019, the fine for

not having health insurance was eliminated.

- Each state is required to have health insurance exchanges, or utilize the federal exchange, whereby insurance companies compete for business. The goal of the exchanges is to improve competition in the market for health insurance.
- Employer mandate: All employers with more than 50 employees must offer health insurance to their employees.

The Affordable Care Act (ACA) is funded through additional taxes that include:

- Increasing the Medicare tax by 0.9 percent and adding a 3.8 percent tax on unearned income for high income taxpayers.
- Charging an annual fee on health insurance providers.
- Imposing other taxes such as a 2.3% tax on manufacturers and importers of certain medical devices.

Many people and politicians, including Donald Trump, have sought to overturn the bill. Those who oppose the bill believe it violates an individual's right to choose whether to have insurance or not. In 2012, a number of states challenged the law on the basis that the individual mandate provision is unconstitutional. In June 2012, the U.S. Supreme Court ruled in a 5–4 decision that the individual mandate is actually a tax, so it is constitutional as the federal government has the right to tax the populace. At the same time, some of the taxes that were implemented as part of the ACA have been eliminated.



BRING IT HOME

What's the Big Deal with Obamacare?

What is it that the Affordable Care Act (ACA) will actually do? To begin with, we should note that it is a massively complex law, with a large number of parts, some of which the Obama administration implemented immediately, and others that the government is supposed to phase in every year from 2013 through 2020. Three of these parts are coverage for the uninsured—those without health insurance, coverage for individuals with preexisting conditions, and the so-called employer and individual mandates, which require employers to offer and people to purchase health insurance. Under the Trump administration, several components of the ACA were repealed or overhauled, while under the Biden administration (and with the support of a majority of the population) the ACA has continued as a major element in provision of health care in the United States.

As we noted in the chapter, people face ever-increasing healthcare costs in the United States. Over the years, the ranks of the uninsured in the United States have grown as rising prices have pushed employers and individuals out of the market. Insurance companies have increasingly used pre-existing medical conditions to determine if someone is high risk, for whom insurance companies either charge higher prices, or they choose to deny insurance coverage to these individuals. Whatever the cause, we noted at the beginning of the chapter that prior to the ACA, more than 32 million Americans were uninsured. People who are uninsured tend to use emergency rooms for treatment—the most expensive form of healthcare, which has contributed significantly to rising costs.

The ACA introduced regulations designed to control increases in healthcare costs. One example is a cap on the amount healthcare providers can spend on administrative costs. Another is a requirement that healthcare providers switch to electronic medical records (EMRs), which will reduce administrative costs.

The ACA required that states establish health insurance exchanges, or markets, where people without health insurance, and businesses that do not provide it for their employees, can shop for different insurance plans. The purpose of these exchanges was to increase competition in insurance markets and thus reduce prices of policies.

Finally, the ACA mandated that people with preexisting conditions could no longer be denied health insurance. The U.S. Department of Health and Human Services estimates that the those without insurance in the US has fallen from 20.3% in 2012 to 11.5% in 2016. Accordingly, 20 million Americans gained coverage under the ACA. According to the Census, as of 2020, the share of the population without health insurance had fallen to 8.6%. So the ACA has

resulted in a decline in the percentage of Americans without health insurance by almost 60%.

What was the cost of this increased coverage and how was it paid? An insurance policy works by insuring against the possibility of needing healthcare. If there are high risk individuals in the insurance pool, the pool must be expanded to include enough low risk individuals to keep average premiums affordable. To that end, the ACA imposed the individual mandate, requiring all individuals to purchase insurance (or pay a penalty) whether they were high risk or not. Many young adults would choose to skip health insurance since the likelihood of their needing significant healthcare is small. The individual mandate brought in a significant amount of money to pay for the ACA. However, despite the elimination of the penalty for not having insurance, ACA coverage has continued to increase. In addition, there were three other funding sources. The ACA took \$716 billion which otherwise would have gone to Medicare spending. The ACA also increased the Medicare tax that wealthy Americans paid by an additional 0.9%. Despite these funding sources, the Congressional Budget Office estimates that the ACA will increase the federal debt by \$137 billion over the next decade.

The impact of the Patient Protection and Affordable Care Act has been a rise in Americans with health insurance. However, due to the increased taxes to pay for the ACA and the increased deficit spending, the ACA faces continued opposition. The Trump administration vowed to repeal it on the campaign trail but no alternative bill has made its way before congress. Only time will tell if the Affordable Care Act will leave a legacy or will quickly be swept by the wayside, jeopardizing the 20 million newly insured Americans.

At the time of this writing, the final impact of the Patient Protection and Affordable Care Act is not clear. Millions of previously uninsured Americans now have coverage, but the increased taxes to pay for ACA and increased deficit spending have created significant political opposition. Whether or not that opposition eventually succeeds in overturning the ACA remains to be seen.

Key Terms

- adverse selection** when groups with inherently higher risks than the average person seek out insurance, thus straining the insurance system
- asymmetric information** a situation where the seller or the buyer has more information than the other regarding the quality of the item for sale
- coinsurance** when an insurance policyholder pays a percentage of a loss, and the insurance company pays the remaining cost
- collateral** something valuable—often property or equipment—that a lender would have a right to seize and sell if the buyer does not repay the loan
- copayment** when an insurance policyholder must pay a small amount for each service, before insurance covers the rest
- cosigner** another person or firm who legally pledges to repay some or all of the money on a loan if the original borrower does not
- deductible** an amount that the insurance policyholders must pay out of their own pocket before the insurance coverage pays anything
- fee-for-service** when medical care providers are paid according to the services they provide
- health maintenance organization (HMO)** an organization that provides health care and is paid a fixed amount per person enrolled in the plan—regardless of how many services are provided
- imperfect information** a situation where either the buyer or the seller, or both, are uncertain about the qualities of what they are buying and selling
- insurance** method of protecting a person from financial loss, whereby policy holders make regular payments to an insurance entity; the insurance firm then remunerates a group member who suffers significant financial damage from an event covered by the policy
- money-back guarantee** a promise that the seller will refund the buyer's money under certain conditions
- moral hazard** when people have insurance against a certain event, they are less likely to guard against that event occurring
- occupational license** licenses issued by government agencies, which indicate that a worker has completed a certain type of education or passed a certain test
- premium** payment made to an insurance company
- risk group** a group that shares roughly the same risks of an adverse event occurring
- service contract** the buyer pays an extra amount and the seller agrees to fix anything specified in the contract that goes wrong for a set time period
- warranty** a promise to fix or replace the good for a certain period of time

Key Concepts and Summary

16.1 The Problem of Imperfect Information and Asymmetric Information

Many make economic transactions in a situation of imperfect information, where either the buyer, the seller, or both are less than 100% certain about the qualities of what they are buying or selling. When information about the quality of products is highly imperfect, it may be difficult for a market to exist.

A “lemon” is a product that turns out, after the purchase, to have low quality. When the seller has more accurate information about the product's quality than the buyer, the buyer will be hesitant to buy, out of fear of purchasing a “lemon.”

Markets have many ways to deal with imperfect information. In goods markets, buyers facing imperfect information about products may depend upon money-back guarantees, warranties, service contracts, and reputation. In labor markets, employers facing imperfect information about potential employees may turn to resumes, recommendations, occupational licenses for certain jobs, and employment for trial periods. In capital markets, lenders facing imperfect information about borrowers may require detailed loan applications and credit checks, cosigners, and collateral.

16.2 Insurance and Imperfect Information

Insurance is a way of sharing risk. People in a group pay premiums for insurance against some unpleasant event, and those in the group who actually experience the unpleasant event then receive some compensation. The fundamental law of insurance is that what the average person pays in over time cannot be less than what the average person gets out. In an actuarially fair insurance policy, the premiums that a person pays to the insurance company are the same as the average amount of benefits for a person in that risk group. Moral hazard arises in insurance markets because those who are insured against a risk will have less reason to take steps to avoid the costs from that risk.

Many insurance policies have deductibles, copayments, or coinsurance. A deductible is the maximum amount that the policyholder must pay out-of-pocket before the insurance company pays the rest of the bill. A copayment is a flat fee that an insurance policy-holder must pay before receiving services. Coinsurance requires the policyholder to pay a certain percentage of costs. Deductibles, copayments, and coinsurance reduce moral hazard by requiring the insured party to bear some of the costs before collecting insurance benefits.

In a fee-for-service health financing system, medical care providers receive reimbursement according to the cost of services they provide. An alternative method of organizing health care is through health maintenance organizations (HMOs), where medical care providers receive reimbursement according to the number of patients they handle, and it is up to the providers to allocate resources between patients who receive more or fewer health care services. Adverse selection arises in insurance markets when insurance buyers know more about the risks they face than does the insurance company. As a result, the insurance company runs the risk that low-risk parties will avoid its insurance because it is too costly for them, while high-risk parties will embrace it because it looks like a good deal to them.

Self-Check Questions

- For each of the following purchases, say whether you would expect the degree of imperfect information to be relatively high or relatively low:
 - Buying apples at a roadside stand
 - Buying dinner at the neighborhood restaurant around the corner
 - Buying a used laptop computer at a garage sale
 - Ordering flowers over the internet for your friend in a different city
- Why is there asymmetric information in the labor market? What signals can an employer look for that might indicate the traits they are seeking in a new employee?
- Why is it difficult to measure health outcomes?

Review Questions

- Why might it be difficult for a buyer and seller to agree on a price when imperfect information exists?
- What do economists (and used-car dealers) mean by a “lemon”?
- What are some ways a seller of goods might reassure a possible buyer who is faced with imperfect information?
- What are some ways a seller of labor (that is, someone looking for a job) might reassure a possible employer who is faced with imperfect information?
- What are some ways that someone looking for a loan might reassure a bank that is faced with imperfect information about whether the borrower will repay the loan?
- What is an insurance premium?

10. In an insurance system, would you expect each person to receive in benefits pretty much what they pay in premiums or is it just that the average benefits paid will equal the average premiums paid?
11. What is an actuarially fair insurance policy?
12. What is the problem of moral hazard?
13. How can moral hazard lead to more costly insurance premiums than one was expected?
14. Define deductibles, copayments, and coinsurance.
15. How can deductibles, copayments, and coinsurance reduce moral hazard?
16. What is the key difference between a fee-for-service healthcare system and a system based on health maintenance organizations?
17. How might adverse selection make it difficult for an insurance market to operate?
18. What are some of the metrics economists use to measure health outcomes?

Critical Thinking Questions

19. You are on the board of directors of a private high school, which is hiring new tenth-grade science teachers. As you think about hiring someone for a job, what are some mechanisms you might use to overcome the problem of imperfect information?
20. A website offers a place for people to buy and sell emeralds, but information about emeralds can be quite imperfect. The website then enacts a rule that all sellers in the market must pay for two independent examinations of their emerald, which are available to the customer for inspection.
 - a. How would you expect this improved information to affect demand for emeralds on this website?
 - b. How would you expect this improved information to affect the quantity of high-quality emeralds sold on the website?
21. How do you think the problem of moral hazard might have affected the safety of sports such as football and boxing when safety regulations started requiring that players wear more padding?
22. To what sorts of customers would an insurance company offer a policy with a high copay? What about a high premium with a lower copay?

Problems

23. Using [Exercise 16.20](#), sketch the effects in parts (a) and (b) on a single supply and demand diagram. What prediction would you make about how the improved information alters the equilibrium quantity and price?
24. Imagine that you can divide 50-year-old men into two groups: those who have a family history of cancer and those who do not. For the purposes of this example, say that 20% of a group of 1,000 men have a family history of cancer, and these men have one chance in 50 of dying in the next year, while the other 80% of men have one chance in 200 of dying in the next year. The insurance company is selling a policy that will pay \$100,000 to the estate of anyone who dies in the next year.
 - a. If the insurance company were selling life insurance separately to each group, what would be the actuarially fair premium for each group?
 - b. If an insurance company were offering life insurance to the entire group, but could not find out about family cancer histories, what would be the actuarially fair premium for the group as a whole?
 - c. What will happen to the insurance company if it tries to charge the actuarially fair premium to the group as a whole rather than to each group separately?



FIGURE 17.1 Building Home Equity Many people choose to purchase their home rather than rent. This chapter explores how the global financial crisis has influenced home ownership. (Credit: “red sold sign” by Diana Parkhouse/ Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- How Businesses Raise Financial Capital
- How Households Supply Financial Capital
- How to Accumulate Personal Wealth

Introduction to Financial Markets



BRING IT HOME

The Housing Bubble and the 2007 Financial Crisis

In 2006, housing equity in the United States peaked at \$13 trillion. That means that the market prices of homes, less what was still owed on the loans they used to buy these houses, equaled \$13 trillion. This was a very good number, since the equity represented the value of the financial asset most U.S. citizens owned.

However, by 2008 this number declined to \$8.8 trillion, and it plummeted further still in 2009. Combined with the decline in value of other financial assets held by U.S. citizens, by 2010, U.S. homeowners' wealth had shrunk \$14 trillion! This is a staggering result, and it affected millions of lives: people had to alter their retirement, housing, and other important consumption decisions. Just about every other large economy in the world suffered a decline in the market value of financial assets, as a result of the 2008-2009 global financial crisis.

This chapter will explain why people purchase houses (other than as a place to live), why they buy other types of

financial assets, and why businesses sell those financial assets in the first place. The chapter will also give us insight into why financial markets and assets go through boom and bust cycles like the one we described here.

When a firm needs to buy new equipment or build a new facility, it often must go to the financial market to raise funds. Usually firms will add capacity during an economic expansion when profits are on the rise and consumer demand is high. Business investment is one of the critical ingredients needed to sustain economic growth. Even in the sluggish 2009 economy, U.S. firms invested \$1.4 trillion in new equipment and structures, in the hope that these investments would generate profits in the years ahead.

Between the end of the recession in 2009 through the second quarter 2013, profits for the S&P 500 companies grew by 9.7% despite the weak economy, with cost cutting and reductions in input costs driving much of that amount, according to the *Wall Street Journal*. [Figure 17.2](#) shows corporate profits after taxes (adjusted for inventory and capital consumption). Despite the steep decline in quarterly net profit in 2008, profits have recovered and surpassed pre-recession levels.

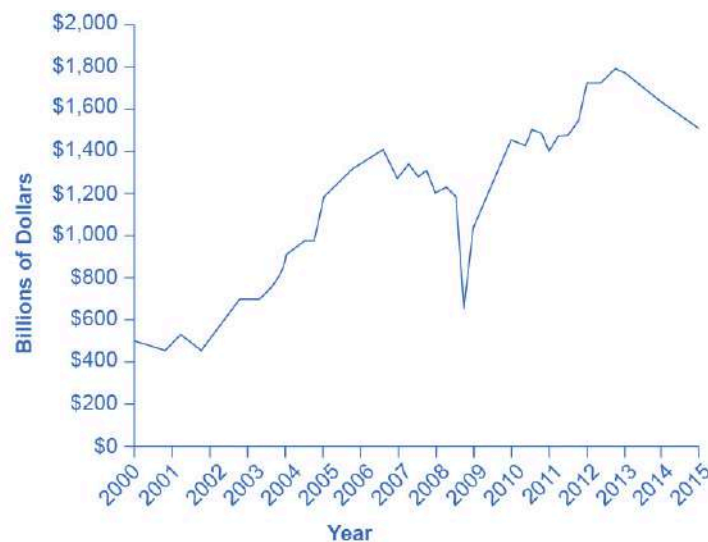


FIGURE 17.2 Corporate Profits After Tax (Adjusted for Inventory and Capital Consumption) Prior to 2008, corporate profits after tax more often than not increased each year. There was a significant drop in profits during 2008 and into 2009. The profit trend has since continued to increase each year, though at a less steady or consistent rate. (Source: Federal Reserve Economic Data (FRED) <https://research.stlouisfed.org/fred2/series/CPATAX>)

Many firms, from huge companies like General Motors to startup firms writing computer software, do not have the financial resources within the firm to make all the desired investments. These firms need financial capital from outside investors, and they are willing to pay interest for the opportunity to obtain a rate of return on the investment of that financial capital.

On the other side of the financial capital market, financial capital suppliers, like households, wish to use their savings in a way that will provide a return. Individuals cannot, however, take the few thousand dollars that they save in any given year, write a letter to General Motors or some other firm, and negotiate to invest their money with that firm. Financial capital markets bridge this gap: that is, they find ways to take the inflow of funds from many separate financial capital suppliers and transform it into the funds of financial capital demanders desire. Such financial markets include stocks, bonds, bank loans, and other financial investments.

[Click to view content \(https://openstax.org/books/principles-economics-3e/pages/17-introduction-to-financial-markets\)](https://openstax.org/books/principles-economics-3e/pages/17-introduction-to-financial-markets)

Corporate Profits After Tax (Adjusted for Inventory and Capital Consumption)

LINK IT UP

Visit this [website \(http://openstax.org/l/marketoverview\)](http://openstax.org/l/marketoverview) to read more about financial markets.

Our perspective then shifts to consider how these financial investments appear to capital suppliers such as the households that are saving funds. Households have a range of investment options: bank accounts, certificates of deposit, money market mutual funds, bonds, stocks, stock and bond mutual funds, housing, and even tangible assets like gold. Finally, the chapter investigates two methods for becoming rich: a quick and easy method that does not work very well at all, and a slow, reliable method that can work very well over a lifetime.

17.1 How Businesses Raise Financial Capital

LEARNING OBJECTIVES

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

- Describe financial capital and how it relates to profits
- Discuss the purpose and process of borrowing, bonds, and corporate stock
- Explain how firms choose between sources of financial capital

Firms often make decisions that involve spending money in the present and expecting to earn profits in the future. Examples include when a firm buys a machine that will last 10 years, or builds a new plant that will last for 30 years, or starts a research and development project. Firms can raise the financial capital they need to pay for such projects in four main ways: (1) from early-stage investors; (2) by reinvesting profits; (3) by borrowing through banks or bonds; and (4) by selling stock. When business owners choose financial capital sources, they also choose how to pay for them.

Early-Stage Financial Capital

Firms that are just beginning often have an idea or a prototype for a product or service to sell, but few customers, or even no customers at all, and thus are not earning profits. Such firms face a difficult problem when it comes to raising financial capital: How can a firm that has not yet demonstrated any ability to earn profits pay a rate of return to financial investors?

For many small businesses, the original source of money is the business owner. Someone who decides to start a restaurant or a gas station, for instance, might cover the startup costs by dipping into their own bank account, or by borrowing money (perhaps using a home as collateral). Alternatively, many cities have a network of well-to-do individuals, known as “angel investors,” who will put their own money into small new companies at an early development stage, in exchange for owning some portion of the firm.

Venture capital firms make financial investments in new companies that are still relatively small in size, but that have potential to grow substantially. These firms gather money from a variety of individual or institutional investors, including banks, institutions like college endowments, insurance companies that hold financial reserves, and corporate pension funds. Venture capital firms do more than just supply money to small startups. They also provide advice on potential products, customers, and key employees. Typically, a venture capital fund invests in a number of firms, and then investors in that fund receive returns according to how the fund as a whole performs.

The amount of money invested in venture capital fluctuates substantially from year to year: as one example, venture capital firms invested more than \$48.3 billion in 2014, according to the National Venture Capital Association. All early-stage investors realize that the majority of small startup businesses will never hit it big; many of them will go out of business within a few months or years. They also know that getting in on the ground floor of a few huge successes like a Netflix or an Amazon.com can make up for multiple failures. Therefore, early-stage investors are willing to take large risks in order to position themselves to gain substantial returns on their investment.

Profits as a Source of Financial Capital

If firms are earning profits (their revenues are greater than costs), they can choose to reinvest some of these profits in equipment, structures, and research and development. For many established companies, reinvesting their own profits is one primary source of financial capital. Companies and firms just getting started may have numerous attractive investment opportunities, but few current profits to invest. Even large firms can experience a year or two of earning low profits or even suffering losses, but unless the firm can find a steady and reliable financial capital source so that it can continue making real investments in tough times, the firm may not survive until better times arrive. Firms often need to find financial capital sources other than profits.

Borrowing: Banks and Bonds

When a firm has a record of at least earning significant revenues, and better still of earning profits, the firm can make a credible promise to pay interest, and so it becomes possible for the firm to borrow money. Firms have two main borrowing methods: banks and bonds.

A bank loan for a firm works in much the same way as a loan for an individual who is buying a car or a house. The firm borrows an amount of money and then promises to repay it, including some rate of interest, over a predetermined period of time. If the firm fails to make its loan payments, the bank (or banks) can often take the firm to court and require it to sell its buildings or equipment to make the loan payments.

Another source of financial capital is a bond. A **bond** is a financial contract: a borrower agrees to repay the amount that it borrowed and also an interest rate over a period of time in the future. A **corporate bond** is issued by firms, but bonds are also issued by various levels of government. For example, a **municipal bond** is issued by cities, a state bond by U.S. states, and a **Treasury bond** by the federal government through the U.S. Department of the Treasury. A bond specifies an amount that one will borrow, the interest rate that one will pay, and the time until repayment.

A large company, for example, might issue bonds for \$10 million. The firm promises to make interest payments at an annual rate of 8%, or \$800,000 per year and then, after 10 years, will repay the \$10 million it originally borrowed. When a firm issues bonds, it may choose to issue many bonds in smaller amounts that together reach the total amount it wishes to raise. A firm that seeks to borrow \$50 million by issuing bonds, might actually issue 10,000 bonds of \$5,000 each. In this way, an individual investor could, in effect, loan the firm \$5,000, or any multiple of that amount. Anyone who owns a bond and receives the interest payments is called a **bondholder**. If a firm issues bonds and fails to make the promised interest payments, the bondholders can take the firm to court and require it to pay, even if the firm needs to raise the money by selling buildings or equipment. However, there is no guarantee the firm will have sufficient assets to pay off the bonds. The bondholders may recoup only a portion of what they loaned the firm.

Bank borrowing is more customized than issuing bonds, so it often works better for relatively small firms. The bank can get to know the firm extremely well—often because the bank can monitor sales and expenses quite accurately by looking at deposits and withdrawals. Relatively large and well-known firms often issue bonds instead. They use bonds to raise new financial capital that pays for investments, or to raise capital to pay off old bonds, or to buy other firms. However, the idea that firms or individuals use banks for relatively smaller loans and bonds for larger loans is not an ironclad rule: sometimes groups of banks make large loans and sometimes relatively small and lesser-known firms issue bonds.

Corporate Stock and Public Firms

A **corporation** is a business that “incorporates”—that is owned by shareholders that have limited liability for the company's debt but share in its profits (and losses). Corporations may be private or public, and may or may not have publicly traded stock. They may raise funds to finance their operations or new investments by raising capital through selling stock or issuing bonds.

Those who buy the stock become the firm's owners, or **shareholders**. **Stock** represents firm ownership; that is,

a person who owns 100% of a company's stock, by definition, owns the entire company. The company's stock is divided into **shares**. Corporate giants like IBM, AT&T, Ford, General Electric, Microsoft, Merck, and Exxon all have millions of stock shares. In most large and well-known firms, no individual owns a majority of the stock shares. Instead, large numbers of shareholders—even those who hold thousands of shares—each have only a small slice of the firm's overall ownership.

When a large number of shareholders own a company, there are three questions to ask:

1. How and when does the company obtain money from its sale of stock?
2. What rate of return does the company promise to pay when it sells stock?
3. Who makes decisions in a company owned by a large number of shareholders?

First, a firm receives money from the stock sale only when the company sells its own stock to the public (the public includes individuals, mutual funds, insurance companies, and pension funds). We call a firm's first stock sale to the public an **initial public offering (IPO)**. The IPO is important for two reasons. For one, the IPO, and any stock issued thereafter, such as stock held as treasury stock (shares that a company keeps in their own treasury) or new stock issued later as a secondary offering, provides the funds to repay the early-stage investors, like the angel investors and the venture capital firms. A venture capital firm may have a 40% ownership in the firm. When the firm sells stock, the venture capital firm sells its part ownership of the firm to the public. A second reason for the importance of the IPO is that it provides the established company with financial capital for substantially expanding its operations.

However, most of the time when one buys and sells corporate stock the firm receives no financial return at all. If you buy General Motors stock, you almost certainly buy it from the current share owner, and General Motors does not receive any of your money. This pattern should not seem particularly odd. After all, if you buy a house, the current owner receives your money, not the original house builder. Similarly, when you buy stock shares, you are buying a small slice of the firm's ownership from the existing owner—and the firm that originally issued the stock is not a part of this transaction.

Second, when a firm decides to issue stock, it must recognize that investors will expect to receive a rate of return. That rate of return can come in two forms. A firm can make a direct payment to its shareholders, called a **dividend**. Alternatively, a financial investor might buy a share of stock in Wal-Mart for \$45 and then later sell it to someone else for \$60, for \$15 gain. We call the increase in the stock value (or of any asset) between when one buys and sells it a **capital gain**.

Third: Who makes the decisions about when a firm will issue stock, or pay dividends, or re-invest profits? To understand the answers to these questions, it is useful to separate firms into two groups: private and public.

A **private company** is frequently owned by the people who generally run it on a day-to-day basis. Individuals can run a private company. We call this a **sole proprietorship**. If a group runs it, we call it a **partnership**. A private company can also be a corporation, but with no publicly issued stock. A small law firm run by one person, even if it employs some other lawyers, would be a sole proprietorship. Partners may jointly own a larger law firm. Most private companies are relatively small, but there are some large private corporations, with tens of billions of dollars in annual sales, that do not have publicly issued stock, such as farm products dealer Cargill, the Mars candy company, and the Bechtel engineering and construction firm.

When a firm decides to sell stock, which financial investors can buy and sell, we call it a **public company**. Shareholders own a public company. Since the shareholders are a very broad group, often consisting of thousands or even millions of investors, the shareholders vote for a board of directors, who in turn hire top executives to run the firm on a day-to-day basis. The more stock a shareholder owns, the more votes that shareholder is entitled to cast for the company's board of directors.

In theory, the board of directors helps to ensure that the firm runs in the interests of the true owners—the shareholders. However, the top executives who run the firm have a strong voice in choosing the candidates

who will serve on their board of directors. After all, few shareholders are knowledgeable enough or have enough personal incentive to spend energy and money nominating alternative board members.

How Firms Choose between Financial Capital Sources

There are clear patterns in how businesses raise financial capital. We can explain these patterns in terms of imperfect information, which as we discussed in [Information, Risk, and Insurance](#), is a situation where buyers and sellers in a market do not both have full and equal information. Those who are actually running a firm will almost always have more information about whether the firm is likely to earn profits in the future than outside investors who provide financial capital.

Any young startup firm is a risk. Some startup firms are only a little more than an idea on paper. The firm's founders inevitably have better information than anyone else about how hard they are willing to work, and whether the firm is likely to succeed. When the founders invested their own money into the firm, they demonstrate a belief in its prospects. At this early stage, angel investors and venture capitalists try to overcome the imperfect information, at least in part, by knowing the managers and their business plan personally and by giving them advice.

Accurate information is sometimes not available because **corporate governance**, the name economists give to the institutions that are supposed to watch over top executives, fails, as the following Clear It Up feature on Lehman Brothers shows.



CLEAR IT UP

How did lack of corporate governance lead to the Lehman Brothers failure?

In 2008, Lehman Brothers was the fourth largest U.S. investment bank, with 25,000 employees. The firm had been in business for 164 years. On September 15, 2008, Lehman Brothers filed for Chapter 11 bankruptcy protection. There are many causes of the Lehman Brothers failure. One area of apparent failure was the lack of oversight by the Board of Directors to keep managers from undertaking excessive risk. We can attribute part of the oversight failure, according to Tim Geithner's April 10, 2010, testimony to Congress, to the Executive Compensation Committee's emphasis on short-term gains without enough consideration of the risks. In addition, according to the court examiner's report, the Lehman Brother's Board of Directors paid too little attention to the details of the operations of Lehman Brothers and also had limited financial service experience.

The board of directors, elected by the shareholders, is supposed to be the first line of corporate governance and oversight for top executives. A second institution of corporate governance is the auditing firm the company hires to review the company's financial records and certify that everything looks reasonable. A third institution of corporate governance is outside investors, especially large shareholders like those who invest large mutual funds or pension funds. In the case of Lehman Brothers, corporate governance failed to provide investors with accurate financial information about the firm's operations.

As a firm becomes at least somewhat established and its strategy appears likely to lead to profits in the near future, knowing the individual managers and their business plans on a personal basis becomes less important, because information has become more widely available regarding the company's products, revenues, costs, and profits. As a result, other outside investors who do not know the managers personally, like bondholders and shareholders, are more willing to provide financial capital to the firm.

At this point, a firm must often choose how to access financial capital. It may choose to borrow from a bank, issue bonds, or issue stock. The great disadvantage of borrowing money from a bank or issuing bonds is that the firm commits to scheduled interest payments, whether or not it has sufficient income. The great advantage of borrowing money is that the firm maintains control of its operations and is not subject to shareholders. Issuing stock involves selling off company ownership to the public and becoming responsible to a board of

directors and the shareholders.

The benefit of issuing stock is that a small and growing firm increases its visibility in the financial markets and can access large amounts of financial capital for expansion, without worrying about repaying this money. If the firm is successful and profitable, the board of directors will need to decide upon a dividend payout or how to reinvest profits to further grow the company. Issuing and placing stock is expensive, requires the expertise of investment bankers and attorneys, and entails compliance with reporting requirements to shareholders and government agencies, such as the federal Securities and Exchange Commission (SEC).

17.2 How Households Supply Financial Capital

LEARNING OBJECTIVES

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

- Show the relationship between savers, banks, and borrowers
- Calculate bond yield
- Contrast bonds, stocks, mutual funds, and assets
- Explain the tradeoffs between return and risk

The ways in which firms would prefer to raise funds are only half the story of financial markets. The other half is what those households and individuals who supply funds desire, and how they perceive the available choices. The focus of our discussion now shifts from firms on the demand side of financial capital markets to households on the supply side of those markets. We can divide the mechanisms for savings available to households into several categories: deposits in bank accounts; bonds; stocks; money market mutual funds; stock and bond mutual funds; and housing and other tangible assets like owning gold. We need to analyze each of these investments in terms of three factors: (1) the expected rate of return it will pay; (2) the risk that the return will be much lower or higher than expected; and (3) the investment's **liquidity**, which refers to how easily one can exchange money or financial assets for a good or service. We will do this analysis as we discuss each of these investments in the sections below. First, however, we need to understand the difference between expected rate of return, risk, and actual rate of return.

Expected Rate of Return, Risk, and Actual Rate of Return

The **expected rate of return** refers to how much a project or an investment is expected to return to the investor, either in future interest payments, capital gains, or increased profitability. It is usually the average return over a period of time, usually in years or even decades. We normally measure it as a percentage rate. **Risk** measures the uncertainty of that project's profitability. There are several types of risk, including default risk and interest rate risk. Default risk, as its name suggests, is the risk that the borrower fails to pay back the bond or loan. Interest rate risk is the danger that you might buy a long term bond at a 6% interest rate right before market rates suddenly rise, so had you waited, you could have received a similar bond that paid 9%. A high-risk investment is one for which a wide range of potential payoffs is reasonably probable. A low-risk investment may have actual returns that are fairly close to its expected rate of return year after year. A high-risk investment will have actual returns that are much higher than the expected rate of return in some months or years and much lower in other months or years. The **actual rate of return** refers to the total rate of return, including capital gains and interest paid on an investment at the end of a time period.

Bank Accounts

An intermediary is one who stands between two other parties. For example, a person who arranges a blind date between two other people is one kind of intermediary. In financial capital markets, banks are an example of a **financial intermediary**—that is, an institution that operates between a saver who deposits funds in a bank and a borrower who receives a loan from that bank. When a bank serves as a financial intermediary, unlike the situation with a couple on a blind date, the saver and the borrower never meet. In fact, it is not even possible to make direct connections between those who deposit funds in banks and those who borrow from banks, because all deposited funds end up in one big pool, which the financial institution then lends out.

Figure 17.3 illustrates the position of banks as a financial intermediary, with a pattern of deposits flowing into a bank and loans flowing out, and then repayment of the loans flowing back to the bank, with interest payments for the original savers.

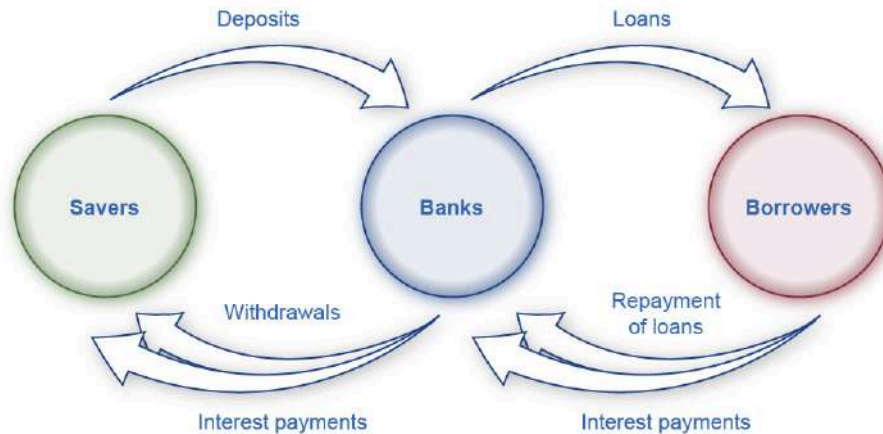


FIGURE 17.3 Banks as Financial Intermediaries Banks are a financial intermediary because they stand between savers and borrowers. Savers place deposits with banks, and then receive interest payments and withdraw money. Borrowers receive loans from banks, and repay the loans with interest.

Banks offer a range of accounts to serve different needs. A **checking account** typically pays little or no interest, but it facilitates transactions by giving you easy access to your money, either by writing a check or by using a **debit card** (that is, a card which works like a credit card, except that purchases are immediately deducted from your checking account rather than billed separately through a credit card company). A **savings account** typically pays some interest rate, but getting the money typically requires you to make a trip to the bank or an automatic teller machine (or you can access the funds electronically). The lines between checking and savings accounts have blurred in the last couple of decades, as many banks offer checking accounts that will pay an interest rate similar to a savings account if you keep a certain minimum amount in the account, or conversely, offer savings accounts that allow you to write at least a few checks per month.

Another way to deposit savings at a bank is to use a **certificate of deposit (CD)**. With a CD, you agree to deposit a certain amount of money, often measured in thousands of dollars, in the account for a stated period of time, typically ranging from a few months to several years. In exchange, the bank agrees to pay a higher interest rate than for a regular savings account. While you can withdraw the money before the allotted time, as the advertisements for CDs always warn, there is “a substantial penalty for early withdrawal.”

Figure 17.4 shows the annual rate of interest paid on a six-month, one-year, and five-year CD since 1984, as reported by Bankrate.com. The interest rates that savings accounts pay are typically a little lower than the CD rate, because financial investors need to receive a slightly higher rate of interest as compensation for promising to leave deposits untouched for a period of time in a CD, and thus forfeiting some liquidity.

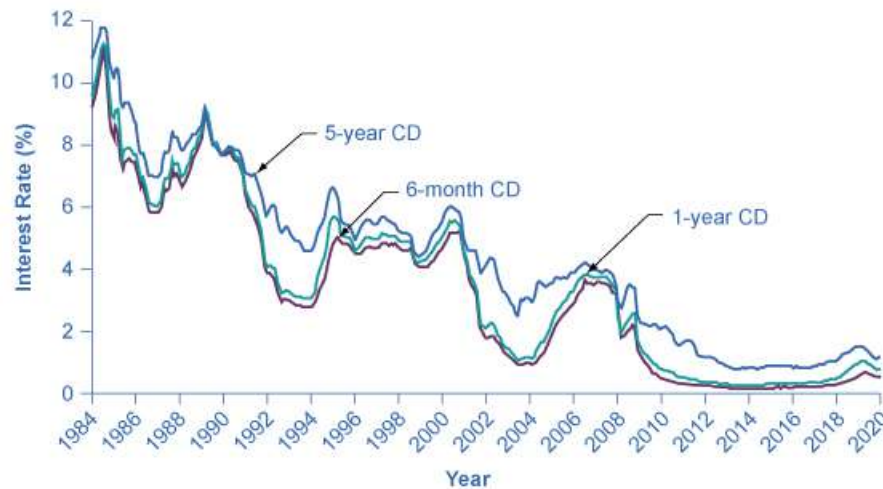


FIGURE 17.4 Interest Rates on Six-Month, One-Year, and Five-Year Certificates of Deposit The interest rates on certificates of deposit have fluctuated over time. The high interest rates of the early 1980s are indicative of the relatively high inflation rate in the United States at that time. Interest rates fluctuate with the business cycle, typically increasing during expansions and decreasing during a recession. Note the steep decline in CD rates since 2008, the beginning of the Great Recession.

The great advantages of bank accounts are that financial investors have very easy access to their money, and also money in bank accounts is extremely safe. In part, this safety arises because a bank account offers more security than keeping a few thousand dollars in the toe of a sock in your underwear drawer. In addition, the Federal Deposit Insurance Corporation (FDIC) protects the savings of the average person. Every bank is required by law to pay a fee to the FDIC, based on the size of its deposits. Then, if a bank should go bankrupt and not be able to repay depositors, the FDIC guarantees that all customers will receive their deposits back up to \$250,000.

The bottom line on bank accounts looks like this: low risk means low rate of return but high liquidity.

Bonds

An investor who buys a bond expects to receive a rate of return. However, bonds vary in the rates of return that they offer, according to the riskiness of the borrower. We always can divide an interest rate into three components (as we explained in [Choice in a World of Scarcity](#)): compensation for delaying consumption, an adjustment for an inflationary rise in the overall level of prices, and a risk premium that takes the borrower's riskiness into account.

The U.S. government is an extremely safe borrower, so when the U.S. government issues Treasury bonds, it can pay a relatively low interest rate. Firms that appear to be safe borrowers, perhaps because of their sheer size or because they have consistently earned profits over time, will pay a higher interest rate than the U.S. government. Firms that appear to be riskier borrowers, perhaps because they are still growing or their businesses appear shaky, will pay the highest interest rates when they issue bonds. We call bonds that offer high interest rates to compensate for their relatively high chance of default **high-yield bonds** or **junk bonds**. A number of today's well-known firms issued junk bonds in the 1980s when they were starting to grow, including Turner Broadcasting and Microsoft.

LINK IT UP

Visit this [website \(http://openstax.org/l/bondsecurities\)](http://openstax.org/l/bondsecurities) to read about Treasury bonds.

A bond issued by the U.S. government or a large corporation may seem to be relatively low risk: after all, the

bond issuer has promised to make certain payments over time, and except for rare bankruptcy cases, these payments will occur. If a corporate bond issuer fails to make the payments that it owes to its bondholders, the bondholders can require that the company declare bankruptcy, sell off its assets, and pay them as much as it can. Even in the case of junk bonds, a wise investor can reduce the risk by purchasing bonds from a wide range of different companies since, even if a few firms go broke and do not pay, they are not all likely to go bankrupt.

As we noted before, bonds carry an interest rate risk. For example, imagine you decide to buy a 10-year bond for \$1,000 that would pay an annual interest rate of 8%. Soon after you buy the bond, interest rates on bonds rise, so that now similar companies are paying an annual rate of 12%. Anyone who buys a \$1,000 bond now can receive annual payments of \$120 per year, but since your bond was issued at an interest rate of 8%, you have tied up \$1,000 and receive payments of only \$80 per year. In the meaningful sense of opportunity cost, you are missing out on the higher payments that you could have received. Furthermore, you can calculate the amount you should be willing to pay now for future payments. To place a present discounted value on a future payment, decide what you would need in the present to equal a certain amount in the future. This calculation will require an interest rate. For example, if the interest rate is 25%, then a payment of \$125 a year from now will have a present discounted value of \$100—that is, you could take \$100 in the present and have \$125 in the future. (We discuss this further in the appendix on [Present Discounted Value](#).)

In financial terms, a bond has several parts. A bond is basically an “I owe you” note that an investor receives in exchange for capital (money). The bond has a **face value**. This is the amount the borrower agrees to pay the investor at maturity. The bond has a **coupon rate** or interest rate, which is usually semi-annual, but can be paid at different times throughout the year. (Bonds used to be paper documents with coupons that investors clipped and turned in to the bank to receive interest.) The bond has a **maturity date** when the borrower will pay back its face value as well as its last interest payment. Combining the bond’s face value, interest rate, and maturity date, and market interest rates, allows a buyer to compute a bond’s **present value**, which is the most that a buyer would be willing to pay for a given bond. This may or may not be the same as the face value.

The **bond yield** measures the rate of return a bond is expected to pay over time. Investors can buy bonds when they are issued and they can buy and sell them during their lifetimes. When buying a bond that has been around for a few years, investors should know that the interest rate printed on a bond is often not the same as the bond yield, even on new bonds. Read the next Work It Out feature to see how this happens.

WORK IT OUT

Calculating the Bond Yield

You have bought a \$1,000 bond whose coupon rate is 8%. To calculate your return or yield, follow these steps:

1. Assume the following:
Face value of a bond: \$1,000
Coupon rate: 8 %
Annual payment: \$80 per year
2. Consider the risk of the bond. If this bond carries no risk, then it would be safe to assume that the bond will sell for \$1,000 when it is issued and pay the purchaser \$80 per year until its maturity, at which time the final interest payment will be made and the original \$1,000 will be repaid. Now, assume that over time the interest rates prevailing in the economy rise to 12% and that there is now only one year left to this bond’s maturity. This makes the bond an unattractive investment, since an investor can find another bond that perhaps pays 12%. To induce the investor to buy the 8% bond, the bond seller will lower its price below its face value of \$1,000.
3. Calculate the bond’s price when its interest rate is less than the market interest rate. The expected payments from the bond one year from now are \$1,080, because in the bond’s last year the bond’s issuer will make the final interest payment and then also repay the original \$1,000. Given that interest rates are

now 12%, you know that you could invest \$964 in an alternative investment and receive \$1,080 a year from now; that is, $\$964(1 + 0.12) = \1080 . Therefore, you will not pay more than \$964 for the original \$1,000 bond.

4. Consider that the investor will receive the \$1,000 face value, plus \$80 for the last year's interest payment. The yield on the bond will be $(\$1080 - \$964)/\$964 = 12\%$. The yield, or total return, means interest payments, plus capital gains. Note that the interest or coupon rate of 8% did not change. When interest rates rise, bonds previously issued at lower interest rates will sell for less than face value. Conversely, when interest rates fall, bonds previously issued at higher interest rates will sell for more than face value.

Figure 17.5 shows bond yield for two kinds of bonds: 10-year Treasury bonds (which are officially called “notes”) and corporate bonds issued by firms that have been given an AAA rating as relatively safe borrowers by Moody's, an independent firm that publishes such ratings. Even though corporate bonds pay a higher interest rate, because firms are riskier borrowers than the federal government, the rates tend to rise and fall together. Treasury bonds typically pay more than bank accounts, and corporate bonds typically pay a higher interest rate than Treasury bonds.

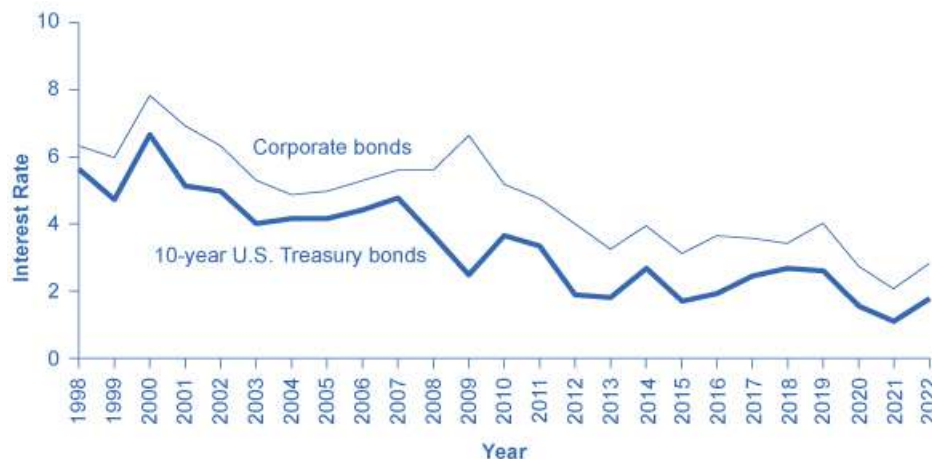


FIGURE 17.5 Interest Rates for Corporate Bonds and Ten-Year U.S. Treasury Bonds The interest rates for corporate bonds and U.S. Treasury bonds (officially “notes”) rise and fall together, depending on conditions for borrowers and lenders in financial markets for borrowing. The corporate bonds always pay a higher interest rate, to make up for the higher risk they have of defaulting compared with the U.S. government.

The bottom line for bonds: rate of return—low to moderate, depending on the borrower's risk; risk—low to moderate, depending on whether interest rates in the economy change substantially after the bond is issued; liquidity—moderate, because the investor needs to sell the bond before the investor regains the cash.

Stocks

As we stated earlier, the rate of return on a financial investment in a share of stock can come in two forms: as dividends paid by the firm and as a capital gain achieved by selling the stock for more than you paid. The range of possible returns from buying stock is mind-bending. Firms can decide to pay dividends or not. A stock price can rise to a multiple of its original price or sink all the way to zero. Even in short periods of time, well-established companies can see large movements in their stock prices. For example, on July 1, 2011, Netflix stock peaked at \$295 per share; one year later, on July 30, 2012, it was at \$53.91 per share; in 2022, it had recovered to \$199. When Facebook went public, its shares of stock sold for around \$40 per share, but in 2022, they were selling for slightly over \$212.

We will discuss the reasons why stock prices fall and rise so abruptly below, but first you need to know how we measure stock market performance. There are a number of different ways to measure the overall performance

of the stock market, based on averaging different subsets of companies' stock prices. Perhaps the best-known stock market measure is the Dow Jones Industrial Average, which is based on 30 large U.S. companies' stock prices. Another stock market performance gauge, the Standard & Poor's 500, follows the stock prices of the 500 largest U.S. companies. The Wilshire 5000 tracks the stock prices of essentially all U.S. companies that have stock the public can buy and sell.

Other stock market measures focus on where stocks are traded. For example, the New York Stock Exchange monitors the performance of stocks that are traded on that exchange in New York City. The Nasdaq stock market includes about 3,600 stocks, with a concentration of technology stocks. [Table 17.1](#) lists some of the most commonly cited measures of U.S. and international stock markets.

Measure of the Stock Market	Comments
Dow Jones Industrial Average (DJIA): https://www.spglobal.com/spdji/en/	Based on 30 large companies from a diverse set of representative industries, chosen by analysts at Dow Jones and Company. The index was started in 1896.
Standard & Poor's 500: http://www.standardandpoors.com	Based on 500 large U.S. firms, chosen by analysts at Standard & Poor's to represent the economy as a whole.
Wilshire 5000: http://www.wilshire.com	Includes essentially all U.S. companies with stock ownership. Despite the name, this index includes about 7,000 firms.
New York Stock Exchange: http://www.nyse.com	The oldest and largest U.S. stock market, dating back to 1792. It trades stocks for 2,800 companies of all sizes. It is located at 18 Broad St. in New York City.
NASDAQ: http://www.nasdaq.com	Founded in 1971 as an electronic stock market, allowing people to buy or sell from many physical locations. It has about 3,600 companies.
FTSE: http://www.ftse.com	Includes the 100 largest companies on the London Stock Exchange. Pronounced "footsie." Originally stood for Financial Times Stock Exchange.
Nikkei: http://www.nikkei.co.jp/nikkeiinfo/en/	Nikkei stands for <i>Nihon Keizai Shimbun</i> , which translates as the Japan Economic Journal, a major business newspaper in Japan. Index includes the 225 largest and most actively traded stocks on the Tokyo Stock Exchange.
DAX: http://www.exchange.de	Tracks 30 of the largest companies on the Frankfurt, Germany, stock exchange. DAX is an abbreviation for <i>Deutscher Aktien Index</i> (German Stock Index).

TABLE 17.1 Some Stock Market Measures

The trend in the stock market is generally up over time, but with some large dips along the way. [Figure 17.6](#) shows the path of the Standard & Poor's 500 index (which is measured on the left-hand vertical axis) and the Dow Jones Index (which is measured on the right-hand vertical axis). Broad stock market measures, like the ones we list here, tend to move together. The S&P 500 Index is the weighted average market capitalization of the firms selected to be in the index. The Dow Jones Industrial Average is the price weighted average of 30 industrial stocks tracked on the New York Stock Exchange.

When the Dow Jones average rises from 5,000 to 10,000, you know that the average price of the stocks in that index has roughly doubled. [Figure 17.6](#) shows that stock prices did not rise much in the 1970s, but then started a steady climb in the 1980s. From 2000 to 2013, stock prices bounced up and down, but ended up at

about the same level.

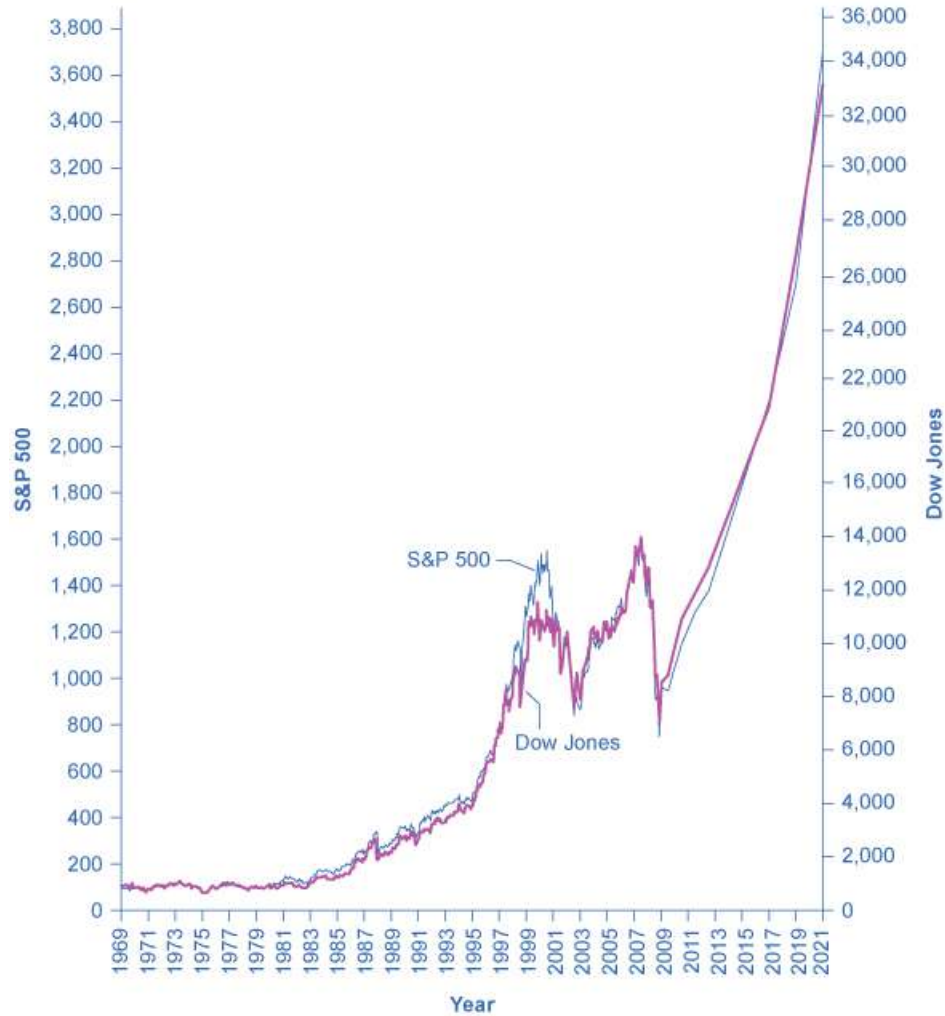


FIGURE 17.6 The Dow Jones Industrial Index and the Standard & Poor's 500, 1965–2021 Stock prices rose dramatically from the 1980s up to about 2000. From 2000 to 2013, stock prices bounced up and down, but ended up at about the same level. Since 2009, both indexes have for the most part increased.

[Table 17.2](#) shows the total annual rate of return an investor would have received from buying the stocks in the S&P 500 index over recent decades. The total return here includes both dividends paid by these companies and also capital gains arising from increases in the stock value. (For technical reasons related to how we calculate the numbers, the dividends and capital gains do not add exactly to the total return.) From the 1950s to the 1980s, the average firm paid annual dividends equal to about 4% of its stock value. Since the 1990s, dividends have dropped and now often provide a return closer to 1% to 2%. In the 1960s and 1970s, the gap between percent earned on capital gains and dividends was much closer than it has been since the 1980s. In the 1980s and 1990s, capital gains were far higher than dividends. In the 2000s, dividends remained low and, while stock prices fluctuated, they ended the decade roughly where they had started. In the 2010s, dividends remained low and stock prices increased, and this continued at the beginning of the 2020s.

Period	Total Annual Return	Capital Gains	Dividends
1950–1959	19.25%	13.58%	4.99%
1960–1969	7.78%	4.39%	3.25%
1970–1979	5.88%	1.60%	4.20%
1980–1989	17.55%	12.59%	4.40%
1990–1999	18.21%	15.31%	2.51%
2000–2009	–1.00%	–2.70%	1.70%
2010–2019	12.65%	10.35%	2.30%
2020	18.40%	16.26%	2.14%
2021	28.71%	26.89%	1.82%

TABLE 17.2 Annual Returns on S&P 500 Stocks, 1950–2021

The overall pattern is that stocks as a group have provided a high rate of return over extended periods of time, but this return comes with risks. The market value of individual companies can rise and fall substantially, both over short time periods and over the long run. During extended periods of time like the 1970s or the first decade of the 2000s, the overall stock market return can be quite modest. The stock market can sometimes fall sharply, as it did in 2008.

The bottom line on investing in stocks is that the rate of return over time will be high, but the risks are also high, especially in the short run. Liquidity is also high since one can sell stock in publicly held companies readily for spendable money.

Mutual Funds

Buying stocks or bonds issued by a single company is always somewhat risky. An individual firm may find itself buffeted by unfavorable supply and demand conditions or hurt by unlucky or unwise managerial decisions. Thus, a standard recommendation from financial investors is **diversification**, which means buying stocks or bonds from a wide range of companies. A saver who diversifies is following the old proverb: “Don’t put all your eggs in one basket.” In any broad group of companies, some firms will do better than expected and some will do worse—but the diversification has a tendency to cancel out extreme increases and decreases in value.

Purchasing a diversified group of stocks or bonds has become easier in the internet age, but it remains something of a task. To simplify the process, companies offer **mutual funds**, which consist of a variety of stocks or bonds from different companies. The financial investor buys mutual fund shares, and then receives a return based on how the fund as a whole performs. In 2021, according to the Investment Company Factbook, just over 47% of U.S. households had a financial investment in a mutual fund—including many people who have their retirement savings or pension money invested in this way.

Mutual funds can focus in certain areas: one mutual fund might invest only in company stocks based in Indonesia, or only in bonds issued by large manufacturing companies, or only in biotechnology companies' stock. At the other end of the spectrum, a mutual fund might be quite broad. At the extreme, some mutual funds own a tiny share of every firm in the stock market, and thus the mutual fund's value will fluctuate with the overall stock market's average. We call a mutual fund that seeks only to mimic the market's overall performance an **index fund**.

Diversification can offset some of the risks of individual stocks rising or falling. Even investors who buy an indexed mutual fund designed to mimic some measure of the broad stock market, like the Standard & Poor's 500, had better prepare against some ups and downs, like those the stock market experienced in the first decade of the 2000s. In 2008 average U.S. stock funds declined 38%, reducing individual and household wealth. This steep drop in value hit hardest those who were close to retirement and were counting on their stock funds to supplement retirement income.

The bottom line on investing in mutual funds is that the rate of return over time will be high. The risks are also high, but the risks and returns for an individual mutual fund will be lower than those for an individual stock. As with stocks, liquidity is also high provided the mutual fund or stock index fund is readily traded.

Housing and Other Tangible Assets

Households can also seek a rate of return by purchasing tangible assets, especially housing. About two-thirds of U.S. households own their own home. An owner's **equity** in a house is the monetary value the owner would have after selling the house and repaying any outstanding bank loans they used to buy the house. For example, imagine that you buy a house for \$200,000, paying 10% of the price as a down payment and taking out a bank loan for the remaining \$180,000. Over time, you pay off some of your bank loan, so that only \$100,000 remains, and the house's value on the market rises to \$250,000. At that point, your equity in the home is the value of the home minus the value of the loan outstanding, which is \$150,000. For many middle-class Americans, home equity is their single greatest financial asset. The total value of all home equity held by U.S. households was \$23.6 trillion as of the middle of 2021, according to Federal Reserve data.

Investment in a house is tangibly different from bank accounts, stocks, and bonds because a house offers both a financial and a nonfinancial return. If you buy a house to live in, part of the return on your investment occurs from your consumption of "housing services"—that is, having a place to live. (Of course, if you buy a home and rent it out, you receive rental payments for the housing services you provide, which would offer a financial return.) Buying a house to live in also offers the possibility of a capital gain from selling the house in the future for more than you paid for it. There can, however, be different outcomes, as the Clear It Up on the housing market shows.

Housing prices have usually risen steadily over time. For example, the median sales price for an existing one-family home was \$122,900 in 1990, but 232,000 at the end of December 2016, according to FRED® Economic Data. Over these 24 years, home prices increased an average of 3.1% per year, which is an average financial return over this time. [Figure 17.7](#) shows U.S. Census data for the average sales price of a new home in the United States from 1965 to 2021.

LINK IT UP

Go to this [website \(http://openstax.org/l/investopedia\)](http://openstax.org/l/investopedia) to experiment with a compound annual growth rate calculator.

However, the possible capital gains from rising housing prices are riskier than these national price averages. Certain regions of the country or metropolitan areas have seen drops in housing prices over time. The median housing price for the United States as a whole fell almost 7% in 2008 and again in 2009, dropping the median price from \$247,900 to \$216,700. As of 2016, home values had recovered and even exceeded their pre-recession levels, and they have continued to increase into the early 2020s.

LINK IT UP

Visit this [website \(http://openstax.org/l/insidejob\)](http://openstax.org/l/insidejob) to watch the trailer for *Inside Job*, a movie that explores the modern financial crisis.

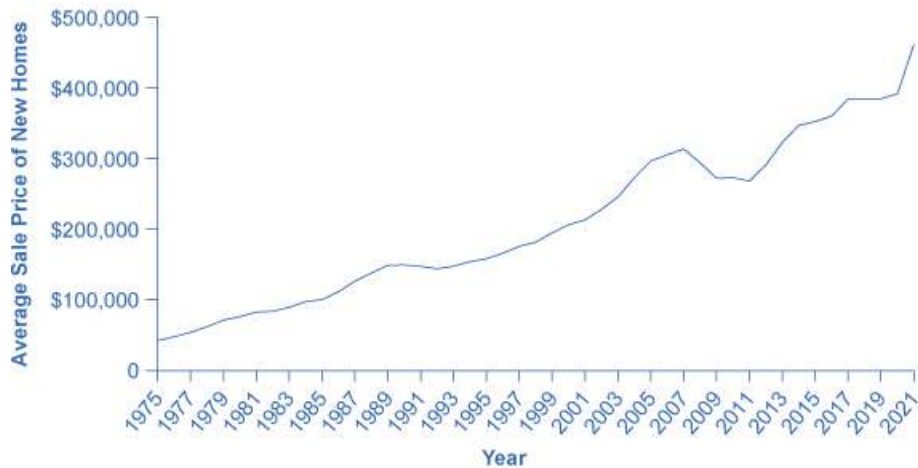


FIGURE 17.7 The Median Average Sales Price for New Single-Family Homes, 1990–2015 The median price is the price where half of sales prices are higher and half are lower. The median sales price for a new one-family home was \$122,900 in 1990. It rose as high as \$248,000 in 2007, before falling to \$232,000 in 2008. In 2015, the median sales price was \$294,000. Of course, this national figure conceals many local differences, like the areas where housing prices are higher or lower, or how housing prices have risen or fallen at certain times. (Source: U.S. Census)

Investors can also put money into other tangible assets such as gold, silver, and other precious metals, or in duller commodities like sugar, cocoa, coffee, orange juice, oil, and natural gas. The return on these investments derives from the saver's hope of buying low, selling high, and receiving a capital gain. Investing in, say, gold or coffee offers relatively little in the way of nonfinancial benefits to the user (unless the investor likes to caress gold or gaze upon a warehouse full of coffee). Typically, investors in these commodities never even see the physical good. Instead, they sign a contract that takes ownership of a certain quantity of these commodities, which are stored in a warehouse, and later they sell the ownership to someone else. As one example, from 1981 to 2005, the gold prices generally fluctuated between about \$300 and \$500 per ounce, but then rose sharply to over \$1,100 per ounce by early 2010. In January 2017, prices were hovering around \$1,191 per ounce, and they have since increased, reaching over \$1,900 by early 2022.

A final area of tangible assets consists of “collectibles” like paintings, fine wine, jewelry, antiques, or even baseball cards. Most collectibles provide returns both in the form of services or of a potentially higher selling price in the future. You can use paintings by hanging them on the wall; jewelry by wearing it; baseball cards by displaying them. You can also hope to sell them someday for more than you paid for them. However, the evidence on prices of collectibles, while scanty, is that while they may go through periods where prices skyrocket for a time, you should not expect to make a higher-than-average rate of return over a sustained period of time from investing in this way.

The bottom line on investing in tangible assets: rate of return—moderate, especially if you can receive nonfinancial benefits from, for example, living in the house; risk—moderate for housing or high if you buy gold or baseball cards; liquidity—low, because it often takes considerable time and energy to sell a house or a piece of fine art and turn your capital gain into cash. The next Clear It Up feature explains the issues in the recent U.S. housing market crisis.



CLEAR IT UP

What was all the commotion in the recent U.S. housing market?

The cumulative average annual growth rate in housing prices from 1981 to 2000 was 5.1%. The price of an average U.S. home then took off from 2003 to 2005, rising more than 10% per year. No serious analyst believed this rate of growth was sustainable; after all, if housing prices grew at, say, 11% per year over time, the average price of a home

would more than double every seven years. However, at the time many serious analysts saw no reason for deep concern. After all, housing prices often change in fits and starts, like all prices, and a price surge for a few years is often followed by prices that are flat or even declining a bit as local markets adjust.

The sharp rise in housing prices was driven by a high level of demand for housing. Interest rates were low, so financial institutions encouraged people to borrow money to buy a house. Banks became much more flexible in their lending, making what were called “subprime” loans. Banks loaned money with low, or sometimes no down payment. They offered loans with very low payments for the first two years, but then much higher payments after that. The idea was that housing prices would keep rising, so the borrower would just refinance the mortgage two years in the future, and thus would not ever have to make the higher payments. Some banks even offered so-called NINJA loans, which meant a financial institution issued a loan even though the borrower had no income, no job, nor assets.

In retrospect, these loans seem outlandish. Many borrowers figured, however, that as long as housing prices kept rising, it made sense to buy. Many lenders used a process called “securitizing,” in which they sold their mortgages to financial companies, which put all the mortgages into a big pool, creating large financial securities, and then re-sold these mortgage-backed securities to investors. In this way, the lenders off-loaded the mortgage risks to investors. Investors were interested in mortgage-backed securities as they appeared to offer a steady stream of income, provided the borrowers repaid them. Investors relied on the ratings agencies to assess the credit risk associated with the mortgage-backed securities. In hindsight, it appears that the credit agencies were far too lenient in their ratings of many of the securitized loans. Bank and financial regulators watched the steady rise in the market for mortgage-backed securities, but saw no reason at the time to intervene.

When housing prices turned down, many households that had borrowed when prices were high found that what they owed the bank was more than what their home was worth. Many banks believed that they had diversified by selling their individual loans and instead buying securities based on mortgage loans from all over the country. After all, banks thought back in 2005, the average house price had not declined at any time since the Great Depression in the 1930s. These securities based on mortgage loans, however, turned out to be far riskier than expected. The bust in housing prices weakened both bank and household finances, and thus helped bring on the 2008–2009 Great Recession.

The Tradeoffs between Return and Risk

The discussion of financial investments has emphasized the expected rate of return, the risk, and the liquidity of each investment. [Table 17.3](#) summarizes these characteristics.

Financial Investment	Return	Risk	Liquidity
Checking account	Very low	Very little	Very high
Savings account	Low	Very little	High
Certificate of deposit	Low to medium	Very little	Medium
Stocks	High	Medium to high	Medium
Bonds	Medium	Low to medium	Medium
Mutual funds	Medium to high	Medium to high	Medium to high
Housing	Medium	Medium	Low

TABLE 17.3 Key Characteristics of Financial Investments

Financial Investment	Return	Risk	Liquidity
Gold	Medium	High	Low
Collectibles	Low to medium	High	Low

TABLE 17.3 Key Characteristics of Financial Investments

The household investment choices listed here display a tradeoff between the expected return and the degree of risk involved. Bank accounts have very low risk and very low returns; bonds have higher risk but higher returns; and stocks are riskiest of all but have the potential for still higher returns. In effect, the higher average return compensates for the higher degree of risk. If risky assets like stocks did not also offer a higher average return, then few investors would want them.

This tradeoff between return and risk complicates the task of any financial investor: Is it better to invest safely or to take a risk and go for the high return? Ultimately, choices about risk and return will be based on personal preferences. However, it is often useful to examine risk and return in the context of different time frames.

The high returns of stock market investments refer to a high average return that we can expect over a period of several years or decades. The high risk of such investments refers to the fact that in shorter time frames, from months to a few years, the rate of return may fluctuate a great deal. Thus, a person near retirement age, who already owns a house, may prefer reduced risk and certainty about retirement income. For young workers, just starting to make a reasonably profitable living, it may make sense to put most of their savings for retirement in mutual funds. Mutual funds are able to take advantage of their buying and selling size and thereby reduce transaction costs for investors. Stocks are risky in the short term, to be sure, but when the worker can look forward to several decades during which stock market ups and downs can even out, stocks will typically pay a much higher return over that extended period than will bonds or bank accounts. Thus, one must consider tradeoffs between risk and return in the context of where the investor is in life.

17.3 How to Accumulate Personal Wealth

LEARNING OBJECTIVES

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

- Explain the random walk theory
- Calculate simple and compound interest
- Evaluate how capital markets transform financial capital

Getting rich may seem straightforward enough. Figure out what companies are going to grow and earn high profits in the future, or figure out what companies are going to become popular for everyone else to buy. Those companies are the ones that will pay high dividends or whose stock price will climb in the future. Then, buy stock in those companies. Presto! Multiply your money!

Why is this path to riches not as easy as it sounds? This module first discusses the problems with picking stocks, and then discusses a more reliable but undeniably duller method of accumulating personal wealth.

Why It Is Hard to Get Rich Quick: The Random Walk Theory

The chief problem with attempting to buy stock in companies that will have higher prices in the future is that many other financial investors are trying to do the same thing. Thus, in attempting to get rich in the stock market, it is no help to identify a company that is going to earn high profits if many other investors have already reached the same conclusion, because the stock price will already be high, based on the expected high level of future profits.

The idea that stock prices are based on expectations about the future has a powerful and unexpected

implication. If expectations determine stock price, then shifts in expectations will determine shifts in the stock price. Thus, what matters for predicting whether the stock price of a company will do well is not whether the company will actually earn profits in the future. Instead, you must find a company that analysts widely believe at present to have poor prospects, but that will actually turn out to be a shining star. Brigades of stock market analysts and individual investors are carrying out such research 24 hours a day.

The fundamental problem with predicting future stock winners is that, by definition, no one can predict the future news that alters expectations about profits. Because stock prices will shift in response to unpredictable future news, these prices will tend to follow what mathematicians call a “random walk with a trend.” The “random walk” part means that, on any given day, stock prices are just as likely to rise as to fall. “With a trend” means that over time, the upward steps tend to be larger than the downward steps, so stocks do gradually climb.

If stocks follow a random walk, then not even financial professionals will be able to choose those that will beat the average consistently. While some investment advisers are better than average in any given year, and some even succeed for a number of years in a row, the majority of financial investors do not outguess the market. If we look back over time, it is typically true that half or two-thirds of the mutual funds that attempted to pick stocks which would rise more than the market average actually ended up performing worse than the market average. For the average investor who reads the newspaper business pages over a cup of coffee in the morning, the odds of doing better than full-time professionals is not very good at all. Trying to pick the stocks that will gain a great deal in the future is a risky and unlikely way to become rich.

Getting Rich the Slow, Boring Way

Many U.S. citizens can accumulate a large amount of wealth during their lifetimes, if they make two key choices. The first is to complete additional education and training. In 2020, the Bureau of Labor Statistics reported median weekly usual earnings for full-time wage and salary workers age 25 and over that corresponded to annual income of \$40,612 for those with a high school diploma, \$48,776 for those with a two-year associate degree, and \$67,860 for those with a four-year bachelor’s degree. Learning is not only good for you, but it pays off financially, too.

The second key choice is to start saving money early in life, and to give the power of compound interest a chance. Imagine that at age 25, you save \$3,000 and place that money into an account that you do not touch. In the long run, it is not unreasonable to assume a 7% real annual rate of return (that is, 7% above the rate of inflation) on money invested in a well-diversified stock portfolio. After 40 years, using the formula for compound interest, the original \$3,000 investment will have multiplied nearly fifteen fold:

$$3,000(1 + .07)^{40} = \$44,923$$

Having \$45,000 does not make you a millionaire. Notice, however, that this tidy sum is the result of saving \$3,000 exactly once. Saving that amount every year for several decades—or saving more as income rises—will multiply the total considerably. This type of wealth will not rival the riches of Microsoft CEO Bill Gates, but remember that only half of Americans have any money in mutual funds at all. Accumulating hundreds of thousands of dollars by retirement is a perfectly achievable goal for a well-educated person who starts saving early in life—and that amount of accumulated wealth will put you at or near the top 10% of all American households. The following Work It Out feature shows the difference between simple and compound interest, and the power of compound interest.

WORK IT OUT

Simple and Compound Interest

Simple interest is an interest rate calculation only on the principal amount.

Step 1. Learn the formula for simple interest:

$$\text{Principal} \times \text{Rate} \times \text{Time} = \text{Interest}$$

Step 2. Practice using the simple interest formula.

Example 1: \$100 Deposit at a simple interest rate of 5% held for one year is:

$$\$100 \times 0.05 \times 1 = \$5$$

Simple interest in this example is \$5.

Example 2: \$100 Deposit at a simple interest rate of 5% held for three years is:

$$\$100 \times 0.05 \times 3 = \$15$$

Simple interest in this example is \$15.

Step 3. Calculate the total future amount using this formula:

$$\text{Total future amount} = \text{principal} + \text{interest}$$

Step 4. Put the two simple interest formulas together.

$$\text{Total future amount (with simple interest)} = \text{Principal} + (\text{Principal} \times \text{Rate} \times \text{Time})$$

Step 5. Apply the simple interest formula to our three year example.

$$\text{Total future amount (with simple interest)} = \$100 + (\$100 \times 0.05 \times 3) = \$115$$

Compound interest is an interest rate calculation on the principal plus the accumulated interest.

Step 6. To find the compound interest, we determine the difference between the future value and the present value of the principal. This is accomplished as follows:

$$\text{Future Value} = \text{Principal} \times (1 + \text{interest rate})^{\text{time}}$$

$$\text{Compound interest} = \text{Future Value} - \text{Present Value}$$

Step 7. Apply this formula to our three-year scenario. Follow the calculations in

[Table 17.4](#)

Year 1	
Amount in Bank	\$100
Bank Interest Rate	5%
Total	\$105
	$\$100 + (\$100 \times 0.05)$
Year 2	
Amount in Bank	\$105
Bank Interest Rate	5%
Total	\$110.25

TABLE 17.4

	$\$105 + (\$105 \times .05)$
Year 3	
Amount in Bank	\$110.25
Bank Interest Rate	5%
Total	\$115.75
	$\$110.25 + (\$110.25 \times .05)$
Compound interest	$\$115.76 - \$100 = \$15.76$

TABLE 17.4

Step 8. Note that, after three years, the total is \$115.76. Therefore the total compound interest is \$15.76. This is \$0.76 more than we obtained with simple interest. While this may not seem like much, keep in mind that we were only working with \$100 and over a relatively short time period. Compound interest can make a huge difference with larger sums of money and over longer periods of time.

Obtaining additional education and saving money early in life obviously will not make you rich overnight. Additional education typically means deferring earning income and living as a student for more years. Saving money often requires choices like driving an older or less expensive car, living in a smaller apartment or buying a smaller house, and making other day-to-day sacrifices. For most people, the tradeoffs for achieving substantial personal wealth will require effort, patience, and sacrifice.

How Capital Markets Transform Financial Flows

Financial capital markets have the power to repackage money as it moves from those who supply financial capital to those who demand it. Banks accept checking account deposits and turn them into long-term loans to companies. Individual firms sell shares of stock and issue bonds to raise capital. Firms make and sell an astonishing array of goods and services, but an investor can receive a return on the company's decisions by buying stock in that company. Financial investors sell and resell stocks and bonds to one another. Venture capitalists and angel investors search for promising small companies. Mutual funds combine the stocks and bonds—and thus, indirectly, the products and investments—of many different companies.

LINK IT UP

Visit this [website \(http://openstax.org/l/austerebaltic/\)](http://openstax.org/l/austerebaltic/) to read an article about how austerity can work. Then visit this [website \(http://openstax.org/l/counteraustere/\)](http://openstax.org/l/counteraustere/) for another perspective on austerity.

In this chapter, we discussed the basic mechanisms of financial markets. (A more advanced course in economics or finance will consider more sophisticated tools.) The fundamentals of those financial capital markets remain the same: Firms are trying to raise financial capital and households are looking for a desirable combination of rate of return, risk, and liquidity. Financial markets are society's mechanisms for bringing together these forces of demand and supply.



BRING IT HOME

The Housing Bubble and the Financial Crisis of 2007

The housing boom and bust in the United States, and the resulting multi-trillion-dollar decline in home equity, began with the fall of home prices starting in 2007. As home values dipped, many home prices fell below the amount the borrower owed on the mortgage and owners stopped paying and defaulted on their loan. Banks found that their assets (loans) became worthless. Many financial institutions around the world had invested in mortgage-backed securities, or had purchased insurance on mortgage-backed securities. When housing prices collapsed, the value of those financial assets collapsed as well. The asset side of the banks' balance sheets dropped, causing bank failures and bank runs. Around the globe, financial institutions were bankrupted or nearly so. The result was a large decrease in lending and borrowing, or a freezing up of available credit. When credit dries up, the economy is on its knees. The crisis was not limited to the United States. Iceland, Ireland, the United Kingdom, Spain, Portugal, and Greece all had similar housing boom and bust cycles, and similar credit freezes.

If businesses cannot access financial capital, they cannot make physical capital investments. Those investments ultimately lead to job creation. When credit dried up, businesses invested less, and they ultimately laid off millions of workers. This caused incomes to drop, which caused demand to drop. In turn businesses sold less, so they laid off more workers. Compounding these events, as economic conditions worsened, financial institutions were even less likely to make loans.

To make matters even worse, as businesses sold less, their expected future profit decreased, and this led to a drop in stock prices. Combining all these effects led to major decreases in incomes, demand, consumption, and employment, and to the Great Recession, which in the United States officially lasted from December 2007 to June 2009. During this time, the unemployment rate rose from 5% to a peak of 10.1%. Four years after the recession officially ended, unemployment was still stubbornly high, at 7.6%, and 11.8 million people were still unemployed.

As the world's leading consumer, if the United States goes into recession, it usually drags other countries down with it. The Great Recession was no exception. With few exceptions, U.S. trading partners also entered into recessions of their own, of varying lengths, or suffered slower economic growth. Like the United States, many European countries also gave direct financial assistance, so-called bailouts, to the institutions that make up their financial markets. There was good reason to do this. Financial markets bridge the gap between demanders and suppliers of financial capital. These institutions and markets need to function in order for an economy to invest in new financial capital.

However, much of this bailout money was borrowed, and this borrowed money contributed to another crisis in Europe. Because of the impact on their budgets of the financial crisis and the resulting bailouts, many countries found themselves with unsustainably high deficits. They chose to undertake austerity measures, large decreases in government spending and large tax increases, in order to reduce their deficits. Greece, Ireland, Spain, and Portugal all had to undertake relatively severe austerity measures. The ramifications of this crisis have spread. Economists even called into question the euro's viability.

Key Terms

actual rate of return the total rate of return, including capital gains and interest paid on an investment at the end of a time period

bond a financial contract through which a borrower like a corporation, a city or state, or the federal government agrees to repay the amount that it borrowed and also a rate of interest over a period of time in the future

bond yield the rate of return a bond is expected to pay at the time of purchase

bondholder someone who owns bonds and receives the interest payments

capital gain a financial gain from buying an asset, like a share of stock or a house, and later selling it at a higher price

certificate of deposit (CD) a mechanism for a saver to deposit funds at a bank and promise to leave them at the bank for a time, in exchange for a higher interest rate

checking account a bank account that typically pays little or no interest, but that gives easy access to money, either by writing a check or by using a “debit card”

compound interest an interest rate calculation on the principal plus the accumulated interest

corporate bond a bond issued by firms that wish to borrow

corporate governance the name economists give to the institutions that are supposed to watch over top executives in companies that shareholders own

corporation a business owned by shareholders who have limited liability for the company’s debt yet a share of the company’s profits; may be private or public and may or may not have publicly-traded stock

coupon rate the interest rate paid on a bond; can be annual or semi-annual

debit card a card that lets the person make purchases, and the financial institution immediately deducts cost from that person’s checking account

diversification investing in a wide range of companies to reduce the level of risk

dividend a direct payment from a firm to its shareholders

equity the monetary value a homeowner would have after selling the house and repaying any outstanding bank loans used to buy the house

expected rate of return how much a project or an investment is expected to return to the investor, either in future interest payments, capital gains, or increased profitability

face value the amount that the bond issuer or borrower agrees to pay the investor

financial intermediary an institution, like a bank, that receives money from savers and provides funds to borrowers

high-yield bonds bonds that offer relatively high interest rates to compensate for their relatively high chance of default

index fund a mutual fund that seeks only to mimic the market’s overall performance

initial public offering (IPO) the first sale of shares of stock by a firm to outside investors

junk bonds see high-yield bonds

liquidity refers to how easily one can exchange money or financial assets for a good or service

maturity date the date that a borrower must repay a bond

municipal bonds a bond issued by cities that wish to borrow

mutual funds funds that buy a range of stocks or bonds from different companies, thus allowing an investor an easy way to diversify

partnership a company run by a group as opposed to an individual

present value a bond’s current price at a given time

private company a firm frequently owned by the people who generally run it on a day-to-day basis

public company a firm that has sold stock to the public, which in turn investors then can buy and sell

risk a measure of the uncertainty of that project’s profitability

savings account a bank account that pays an interest rate, but withdrawing money typically requires a trip to the bank or an automatic teller machine

shareholders people who own at least some shares of stock in a firm

shares a firm's stock, divided into individual portions

simple interest an interest rate calculation only on the principal amount

sole proprietorship a company run by an individual as opposed to a group

stock a specific firm's claim on partial ownership

Treasury bond a bond issued by the federal government through the U.S. Department of the Treasury

venture capital financial investments in new companies that are still relatively small in size, but that have potential to grow substantially

Key Concepts and Summary

17.1 How Businesses Raise Financial Capital

Companies can raise early-stage financial capital in several ways: from their owners' or managers' personal savings, or credit cards and from private investors like angel investors and venture capital firms.

A bond is a financial contract through which a borrower agrees to repay the amount that it borrowed. A bond specifies an amount that one will borrow, the amounts that one will repay over time based on the interest rate when the bond is issued, and the time until repayment. Corporate bonds are issued by firms; municipal bonds are issued by cities, state bonds by U.S. states, and Treasury bonds by the federal government through the U.S. Department of the Treasury.

Stock represents firm ownership. A company's stock is divided into shares. A firm receives financial capital when it sells stock to the public. We call a company's first stock sale to the public the initial public offering (IPO). However, a firm does not receive any funds when one shareholder sells stock in the firm to another investor. One receives the rate of return on stock in two forms: dividends and capital gains.

A private company is usually owned by the people who run it on a day-to-day basis, although hired managers can run it. We call a private company owned and run by an individual a sole proprietorship, while a firm owned and run by a group is a partnership. When a firm decides to sell stock that financial investors can buy and sell, then the firm is owned by its shareholders—who in turn elect a board of directors to hire top day-to-day management. We call this a public company. Corporate governance is the name economists give to the institutions that are supposed to watch over top executives, though it does not always work.

17.2 How Households Supply Financial Capital

We can categorize all investments according to three key characteristics: average expected return, degree of risk, and liquidity. To obtain a higher rate of return, an investor must typically accept either more risk or less liquidity. Banks are an example of a financial intermediary, an institution that operates to coordinate supply and demand in the financial capital market. Banks offer a range of accounts, including checking accounts, savings accounts, and certificates of deposit. Under the Federal Deposit Insurance Corporation (FDIC), banks purchase insurance against the risk of a bank failure.

A typical bond promises the financial investor a series of payments over time, based on the interest rate at the time the financial institution issues the bond, and when the borrower repays it. Bonds that offer a high rate of return but also a relatively high chance of defaulting on the payments are called high-yield or junk bonds. The bond yield is the rate of return that a bond promises to pay at the time of purchase. Even when bonds make payments based on a fixed interest rate, they are somewhat risky, because if interest rates rise for the economy as a whole, an investor who owns bonds issued at lower interest rates is now locked into the low rate and suffers a loss.

Changes in the stock price depend on changes in expectations about future profits. Investing in any individual firm is somewhat risky, so investors are wise to practice diversification, which means investing in a range of companies. A mutual fund purchases an array of stocks and/or bonds. An investor in the mutual fund then receives a return depending on the fund's overall performance as a whole. A mutual fund that seeks to imitate

the overall behavior of the stock market is called an index fund.

We can also regard housing and other tangible assets as forms of financial investment, which pay a rate of return in the form of capital gains. Housing can also offer a nonfinancial return—specifically, you can live in it.

17.3 How to Accumulate Personal Wealth

It is extremely difficult, even for financial professionals, to predict changes in future expectations and thus to choose the stocks whose price will rise in the future. Most Americans can accumulate considerable financial wealth if they follow two rules: complete significant additional education and training after graduating from high school and start saving money early in life.

Self-Check Questions

1. Answer these three questions about early-stage corporate finance:
 - a. Why do very small companies tend to raise money from private investors instead of through an IPO?
 - b. Why do small, young companies often prefer an IPO to borrowing from a bank or issuing bonds?
 - c. Who has better information about whether a small firm is likely to earn profits, a venture capitalist or a potential bondholder, and why?
2. From a firm's point of view, how is a bond similar to a bank loan? How are they different?
3. Calculate the equity each of these people has in their home:
 - a. Eva just bought a house for \$200,000 by putting 10% as a down payment and borrowing the rest from the bank.
 - b. Freda bought a house for \$150,000 in cash, but if she were to sell it now, it would sell for \$250,000.
 - c. Ben bought a house for \$100,000. He put 20% down and borrowed the rest from the bank. However, the value of the house has now increased to \$160,000 and he has paid off \$20,000 of the bank loan.
4. Which has a higher average return over time: stocks, bonds, or a savings account? Explain your answer.
5. Investors sometimes fear that a high-risk investment is especially likely to have low returns. Is this fear true? Does a high risk mean the return must be low?
6. What is the total amount of interest from a \$5,000 loan after three years with a simple interest rate of 6%?
7. If you receive \$500 in simple interest on a loan that you made for \$10,000 for five years, what was the interest rate you charged?
8. You open a 5-year CD for \$1,000 that pays 2% interest, compounded annually. What is the value of that CD at the end of the five years?

Review Questions

9. What are the most common ways for start-up firms to raise financial capital?
10. Why can firms not just use their own profits for financial capital, with no need for outside investors?
11. Why are banks more willing to lend to well-established firms?
12. What is a bond?
13. What does a share of stock represent?
14. When do firms receive money from a stock sale in their firm and when do they not receive money?
15. What is a dividend?
16. What is a capital gain?

17. What is the difference between a private company and a public company?
18. How do the shareholders who own a company choose the actual company managers?
19. Why are banks called “financial intermediaries”?
20. Name several different kinds of bank account. How are they different?
21. Why are bonds somewhat risky to buy, even though they make predetermined payments based on a fixed rate of interest?
22. Why should a financial investor care about diversification?
23. What is a mutual fund?
24. What is an index fund?
25. How is buying a house to live in a type of financial investment?
26. Why is it hard to forecast future movements in stock prices?
27. What are the two key choices U.S. citizens need to make that determines their relative wealth?
28. Is investing in housing always a very safe investment?

Critical Thinking Questions

29. If you owned a small firm that had become somewhat established, but you needed a surge of financial capital to carry out a major expansion, would you prefer to raise the funds through borrowing or by issuing stock? Explain your choice.
30. Explain how a company can fail when the safeguards that should be in place fail.
31. What are some reasons why the investment strategy of a 30-year-old might differ from the investment strategy of a 65-year-old?
32. Explain why a financial investor in stocks cannot earn high capital gains simply by buying companies with a demonstrated record of high profits.
33. Explain what happens in an economy when the financial markets limit access to capital. How does this affect economic growth and employment?
34. You and your friend have opened an account on E-Trade and have each decided to select five similar companies in which to invest. You are diligent in monitoring your selections, tracking prices, current events, and actions the company has taken. Your friend chooses his companies randomly, pays no attention to the financial news, and spends his leisure time focused on everything besides his investments. Explain what might be the performance for each of your portfolios at the end of the year.
35. How do bank failures cause the economy to go into recession?

Problems

36. The Darkroom Windowshade Company has 100,000 shares of stock outstanding. The investors in the firm own the following numbers of shares: investor 1 has 20,000 shares; investor 2 has 18,000 shares; investor 3 has 15,000 shares; investor 4 has 10,000 shares; investor 5 has 7,000 shares; and investors 6 through 11 have 5,000 shares each. What is the minimum number of investors it would take to vote to change the company's top management? If investors 1 and 2 agree to vote together, can they be certain of always getting their way in how the company will be run?

- 37.** Imagine that a local water company issued \$10,000 ten-year bond at an interest rate of 6%. You are thinking about buying this bond one year before the end of the ten years, but interest rates are now 9%.
- Given the change in interest rates, would you expect to pay more or less than \$10,000 for the bond?
 - Calculate what you would actually be willing to pay for this bond.
- 38.** Suppose Ford Motor Company issues a five year bond with a face value of \$5,000 that pays an annual coupon payment of \$150.
- What is the interest rate Ford is paying on the borrowed funds?
 - Suppose the market interest rate rises from 3% to 4% a year after Ford issues the bonds. Will the value of the bond increase or decrease?
- 39.** How much money do you have to put into a bank account that pays 10% interest compounded annually to have \$10,000 in ten years?
- 40.** Many retirement funds charge an administrative fee each year equal to 0.25% on managed assets. Suppose that Alexx and Spenser each invest \$5,000 in the same stock this year. Alexx invests directly and earns 5% a year. Spenser uses a retirement fund and earns 4.75%. After 30 years, how much more will Alexx have than Spenser?



FIGURE 18.1 Domestic Tires? While these tires may all appear similar, some are made in the United States and others are not. Those that are not could be subject to a tariff that could cause the cost of all tires to be higher. (Credit: "Tires" by Jayme del Rosario/Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Voter Participation and Costs of Elections
- Special Interest Politics
- Flaws in the Democratic System of Government

Introduction to Public Economy



BRING IT HOME

Chinese Tire Tariffs

Do you know where the tires on your car are made? If they were imported, they may be subject to a tariff (a tax on imported goods) that could raise the price of your car. What do you think about that tariff? Would you write to your representative or your senator about it? Would you start a Facebook or Twitter campaign?

Most people are unlikely to fight this kind of tax or even inform themselves about the issue in the first place. In *The Logic of Collective Action* (1965), economist Mancur Olson challenged the popular idea that, in a democracy, the majority view will prevail, and in doing so launched the modern study of public economy, sometimes referred to as public choice, a subtopic of microeconomics. In this chapter, we will look at the economics of government policy, why smaller, more organized groups have an incentive to work hard to enact certain policies, and why lawmakers

ultimately make decisions that may result in bad economic policy.

As President Abraham Lincoln famously said in his 1863 *Gettysburg Address*, democratic governments are supposed to be “of the people, by the people, and for the people.” Can we rely on democratic governments to enact sensible economic policies? After all, they react to voters, not to analyses of demand and supply curves. The main focus of an economics course is, naturally enough, to analyze the characteristics of markets and purely economic institutions. However, political institutions also play a role in allocating society’s scarce resources, and economists have played an active role, along with other social scientists, in analyzing how such political institutions work.

Other chapters of this book discuss situations in which market forces can sometimes lead to undesirable results: monopoly, imperfect competition, and antitrust policy; negative and positive externalities; poverty and inequality of incomes; failures to provide insurance; and financial markets that may go from boom to bust. Many of these chapters suggest that the government’s economic policies could address these issues.

However, just as markets can face issues and problems that lead to undesirable outcomes, a democratic system of government can also make mistakes, either by enacting policies that do not benefit society as a whole or by failing to enact policies that would have benefited society as a whole. This chapter discusses some practical difficulties of democracy from an economic point of view: we presume the actors in the political system follow their own self-interest, which is not necessarily the same as the public good. For example, many of those who are eligible to vote do not, which obviously raises questions about whether a democratic system will reflect everyone’s interests. Benefits or costs of government action are sometimes concentrated on small groups, which in some cases may organize and have a disproportionately large impact on politics and in other cases may fail to organize and end up neglected. A legislator who worries about support from voters in their district may focus on spending projects specific to the district without sufficient concern for whether this spending is in the nation’s interest.

When more than two choices exist, the principle that the majority of voters should decide may not always make logical sense, because situations can arise where it becomes literally impossible to decide what the “majority” prefers. Government may also be slower than private firms to correct its mistakes, because government agencies do not face competition or the threat of new entry.

18.1 Voter Participation and Costs of Elections

LEARNING OBJECTIVES

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

- Explain the significance of rational ignorance
- Evaluate the impact of election expenses

In U.S. presidential elections over the last few decades, about 55% to 65% of voting-age citizens actually voted, according to the U.S. Census. In congressional elections when there is no presidential race, or in local elections, the turnout is typically lower, often less than half the eligible voters. In other countries, the share of adults who vote is often higher. For example, in national elections since the 1980s in Germany, Spain, and France, about 75% to 80% of those of voting age cast ballots. Even this total falls well short of 100%. Some countries have laws that require voting, among them Australia, Belgium, Italy, Greece, Turkey, Singapore, and most Latin American nations. At the time the United States was founded, voting was mandatory in Virginia, Maryland, Delaware, and Georgia. Even if the law can require people to vote, however, no law can require that each voter cast an informed or a thoughtful vote. Moreover, in the United States and in most countries around the world, the freedom to vote has also typically meant the freedom *not* to vote.

Why do people not vote? Perhaps they do not care too much about who wins, or they are uninformed about who is running, or they do not believe their vote will matter or change their lives in any way. These reasons are probably tied together, since people who do not believe their vote matters will not bother to become informed

or care who wins. Economists have suggested why a utility-maximizing person might rationally decide not to vote or not to become informed about the election. While a single vote may decide a few elections in very small towns, in most elections of any size, the Board of Elections measures the margin of victory in hundreds, thousands, or even millions of votes. A rational voter will recognize that one vote is extremely unlikely to make a difference. This theory of **rational ignorance** holds that people will not vote if the costs of becoming informed and voting are too high, or they feel their vote will not be decisive in the election.

In a 1957 work, *An Economic Theory of Democracy*, the economist Anthony Downs stated the problem this way: “It seems probable that for a great many citizens in a democracy, rational behavior excludes any investment whatever in political information per se. No matter how significant a difference between parties is revealed to the rational citizen by his free information, or how uncertain he is about which party to support, he realizes that his vote has almost no chance of influencing the outcome... He will not even utilize all the free information available, since assimilating it takes time.” In his classic 1948 novel *Walden Two*, the psychologist B. F. Skinner puts the issue even more succinctly via one of his characters, who states: “The chance that one man’s vote will decide the issue in a national election...is less than the chance that he will be killed on his way to the polls.” The following Clear It Up feature explores another aspect of the election process: spending.



CLEAR IT UP

How much is too much to spend on an election?

In the 2020 elections, it is estimated that spending for president, Congress, and state and local offices amounted to \$14.4 billion, more than twice what had been spent in 2016. The money raised went to the campaigns, including advertising, fundraising, travel, and staff. Many people worry that politicians spend too much time raising money and end up entangled with special interest groups that make major donations. Critics would prefer a system that restricts what candidates can spend, perhaps in exchange for limited public campaign financing or free television advertising time.

How much spending on campaigns is too much? Five billion dollars will buy many potato chips, but in the U.S. economy, which was nearly \$21 trillion in 2020, the \$14.4 billion spent on political campaigns was about 1/15th of 1% of the overall economy. Here is another way to think about campaign spending. *Total* government spending programs in 2020, including federal and state governments, was about \$8.8 trillion, so the cost of choosing the people who would determine how to spend this money was less than 2/10 of 1% of that. In the context of the enormous U.S. economy, \$14.4 billion is not as much money as it sounds. U.S. consumers spend almost \$2 billion per year on toothpaste and \$7 billion on hair care products. In 2020, Proctor and Gamble spent almost \$5 billion on advertising. It may seem peculiar that one company’s spending on advertisements amounts to one third of what is spent on presidential and other elections.

Whatever we believe about whether candidates and their parties spend too much or too little on elections, the U.S. Supreme Court has placed limits on how government can limit campaign spending. In a 1976 decision, *Buckley v. Valeo*, the Supreme Court emphasized that the First Amendment to the U.S. Constitution specifies freedom of speech. The federal government and states can offer candidates a voluntary deal in which government makes some public financing available to candidates, but only if the candidates agree to abide by certain spending limits. Of course, candidates can also voluntarily agree to set certain spending limits if they wish. However, government cannot forbid people or organizations to raise and spend money above these limits if they choose.

In 2002, Congress passed and President George W. Bush signed into law the Bipartisan Campaign Reform Act (BCRA). The relatively noncontroversial portions of the act strengthen the rules requiring full and speedy disclosure of who contributes money to campaigns. However, some controversial portions of the Act limit the ability of individuals and groups to make certain kinds of political donations and they ban certain kinds of advertising in the months leading up to an election. Some called these bans into question after the release of two films: Michael Moore’s *Fahrenheit 9/11* and Citizens United’s *Hillary: The Movie*. At question was whether each film sought to

discredit political candidates for office too close to an election, in violation of the BCRA. The lower courts found that Moore's film did not violate the Act, while Citizens United's did. The fight reached the Supreme Court, as *Citizens United v. Federal Election Commission*, saying that the First Amendment protects the rights of corporations as well as individuals to donate to political campaigns. The Court ruled, in a 5–4 decision, that the spending limits were unconstitutional. This controversial decision, which essentially allows unlimited contributions by corporations to political action committees, overruled several previous decisions and will likely be revisited in the future, due to the strength of the public reaction. For now, it has resulted in a sharp increase in election spending.

While many U.S. adults do not bother to vote in presidential elections, more than half do. What motivates them? Research on voting behavior has indicated that people who are more settled or more “connected” to society tend to vote more frequently. According to the *Washington Post*, more married people vote than single people. Those with a job vote more than the unemployed. Those who have lived longer in a neighborhood are more likely to vote than newcomers. Those who report that they know their neighbors and talk to them are more likely to vote than socially isolated people. Those with a higher income and level of education are also more likely to vote. These factors suggest that politicians are likely to focus more on the interests of married, employed, well-educated people with at least a middle-class level of income than on the interests of other groups. For example, those who vote may tend to be more supportive of financial assistance for the two-year and four-year colleges they expect their children to attend than they are of medical care or public school education aimed at families of unemployed people and those experiencing poverty.

LINK IT UP

Visit this [website \(http://openstax.org/l/votergroups\)](http://openstax.org/l/votergroups) to see a breakdown of how different groups voted in 2020.

There have been many proposals to encourage greater voter turnout: making it easier to register to vote, keeping the polls open for more hours, or even moving Election Day to the weekend, when fewer people need to worry about jobs or school commitments. However, such changes do not seem to have caused a long-term upward trend in the number of people voting. After all, casting an informed vote will always impose some costs of time and energy. It is not clear how to strengthen people's feeling of connectedness to society in a way that will lead to a substantial increase in voter turnout. Without greater voter turnout, however, politicians elected by the votes of 60% or fewer of the population may not enact economic policy in the best interests of 100% of the population. Meanwhile, countering a long trend toward making voting easier, many states have recently enacted new voting laws that critics say are actually barriers to voting. States have passed laws reducing early voting, restricting groups who are organizing get-out-the-vote efforts, enacted strict photo ID laws, as well as laws that require showing proof of U.S. citizenship. The ACLU argues that while these laws profess to prevent voter fraud, they are in effect making it harder for individuals to cast their vote.

18.2 Special Interest Politics

LEARNING OBJECTIVES

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

- Explain how special interest groups and lobbyists can influence campaigns and elections
- Describe pork-barrel spending and logrolling

Many political issues are of intense interest to a relatively small group, as we noted above. For example, many U.S. drivers do not much care where their car tires were made—they just want good quality as inexpensively as possible. In September 2009, President Obama and Congress enacted a tariff (taxes added on imported goods) on tires imported from China that would increase the price by 35 percent in its first year, 30 percent in its second year, and 25 percent in its third year. Interestingly, the U.S. companies that make tires did not favor this step, because most of them also import tires from China and other countries. (See [Globalization and Protectionism](#) for more on tariffs.) However, the United Steelworkers union, which had seen jobs in the tire

industry fall by 5,000 over the previous five years, lobbied fiercely for the tariff. With this tariff, the cost of all tires increased significantly. (See the closing Bring It Home feature at the end of this chapter for more information on the tire tariff.)

Special interest groups are groups that are small in number relative to the nation, but quite well organized and focused on a specific issue. A special interest group can pressure legislators to enact public policies that do not benefit society as a whole. Imagine an environmental rule to reduce air pollution that will cost 10 large companies \$8 million each, for a total cost of \$80 million. The social benefits from enacting this rule provide an average benefit of \$10 for every person in the United States, for a total of about \$3 trillion. Even though the benefits are far higher than the costs for society as a whole, the 10 companies are likely to lobby much more fiercely to avoid \$8 million in costs than the average person is to argue for \$10 worth of benefits.

As this example suggests, we can relate the problem of special interests in politics to an issue we raised in [Environmental Protection and Negative Externalities](#) about economic policy with respect to negative externalities and pollution—the problem called regulatory capture (which we defined in [Monopoly and Antitrust Policy](#)). In legislative bodies and agencies that write laws and regulations about how much corporations will pay in taxes, or rules for safety in the workplace, or instructions on how to satisfy environmental regulations, you can be sure the specific industry affected has lobbyists who study every word and every comma. They talk with the legislators who are writing the legislation and suggest alternative wording. They contribute to the campaigns of legislators on the key committees—and may even offer those legislators high-paying jobs after they have left office. As a result, it often turns out that those regulated can exercise considerable influence over the regulators.

LINK IT UP

Visit this [website \(http://openstax.org/l/lobbying\)](http://openstax.org/l/lobbying) to read about lobbying.

In the early 2000s, about 40 million people in the United States were eligible for Medicare, a government program that provides health insurance for those 65 and older. On some issues, the elderly are a powerful interest group. They donate money and time to political campaigns, and in the 2020 presidential election, 76% of those ages 65–74 voted, while just 51% of those aged 18 to 24 cast a ballot, according to the U.S. Census.

In 2003, Congress passed and President George Bush signed into law a substantial expansion of Medicare that helped the elderly to pay for prescription drugs. The prescription drug benefit cost the federal government about \$40 billion in 2006, and the Medicare system projected that the annual cost would rise to \$121 billion by 2016. The political pressure to pass a prescription drug benefit for Medicare was apparently quite high, while the political pressure to assist the 40 million with no health insurance at all was considerably lower. One reason might be that the American Association for Retired People AARP, a well-funded and well-organized lobbying group represents senior citizens, while there is no umbrella organization to lobby for those without health insurance.

In the battle over passage of the 2010 Affordable Care Act (ACA), which became known as “Obamacare,” there was heavy lobbying on all sides by insurance companies and pharmaceutical companies. However, labor unions and community groups financed a lobby group, Health Care for America Now (HCAN), to offset corporate lobbying. HCAN, spending \$60 million dollars, was successful in helping pass legislation which added new regulations on insurance companies and a mandate that all individuals will obtain health insurance by 2014. The following Work It Out feature further explains voter incentives and lobbyist influence.

WORK IT OUT

Paying To Get Your Way

Suppose Congress proposes a tax on carbon emissions for certain factories in a small town of 10,000 people. Congress estimates the tax will reduce pollution to such an extent that it will benefit each resident by an equivalent of \$300. The tax will also reduce profits to the town's two large factories by \$1 million each. How much should the factory owners be willing to spend to fight the tax passage, and how much should the townspeople be willing to pay to support it? Why is society unlikely to achieve the optimal outcome?

Step 1. The two factory owners each stand to lose \$1 million if the tax passes, so each should be willing to spend up to that amount to prevent the passage, a combined sum of \$2 million. Of course, in the real world, there is no guarantee that lobbying efforts will be successful, so the factory owners may choose to invest an amount that is substantially lower.

Step 2. There are 10,000 townspeople, each standing to benefit by \$300 if the tax passes. Theoretically, then, they should be willing to spend up to \$3 million ($10,000 \times \300) to ensure passage. (Again, in the real world with no guarantees of success, they may choose to spend less.)

Step 3. It is costly and difficult for 10,000 people to coordinate in such a way as to influence public policy. Since each person stands to gain only \$300, many may feel lobbying is not worth the effort.

Step 4. The two factory owners, however, find it very easy and profitable to coordinate their activities, so they have a greater incentive to do so.

Special interests may develop a close relationship with one political party, so their ability to influence legislation rises and falls as that party moves in or out of power. A special interest may even hurt a political party if it appears to a number of voters that the relationship is too cozy. In a close election, a small group that has been under-represented in the past may find that it can tip the election one way or another—so that group will suddenly receive considerable attention. Democratic institutions produce an ebb and flow of political parties and interests and thus offer both opportunities for special interests and ways of counterbalancing those interests over time.

Identifiable Winners, Anonymous Losers

A number of economic policies produce gains whose beneficiaries are easily identifiable, but costs that are partly or entirely shared by a large number who remain anonymous. A democratic political system probably has a bias toward those who are identifiable.

For example, policies that impose price controls—like rent control—may look as if they benefit renters and impose costs only on landlords. However, when landlords then decide to reduce the number of rental units available in the area, a number of people who would have liked to rent an apartment end up living somewhere else because no units were available. These would-be renters have experienced a cost of rent control, but it is hard to identify who they are.

Similarly, policies that block imports will benefit the firms that would have competed with those imports—and workers at those firms—who are likely to be quite visible. Consumers who would have preferred to purchase the imported products, and who thus bear some costs of the protectionist policy, are much less visible.

Specific tax breaks and spending programs also have identifiable winners and impose costs on others who are hard to identify. Special interests are more likely to arise from a group that is easily identifiable, rather than from a group where some of those who suffer may not even recognize they are bearing costs.

Pork Barrels and Logrolling

Politicians have an incentive to ensure that they spend government money in their home state or district, where it will benefit their constituents in a direct and obvious way. Thus, when legislators are negotiating over whether to support a piece of legislation, they commonly ask each other to include **pork-barrel spending**, legislation that benefits mainly a single political district. Pork-barrel spending is another case in which concentrated benefits and widely dispersed costs challenge democracy: the benefits of pork-barrel spending are obvious and direct to local voters, while the costs are spread over the entire country. Read the following Clear It Up feature for more information on pork-barrel spending.



CLEAR IT UP

How much impact can pork-barrel spending have?

Many observers widely regard U.S. Senator Robert C. Byrd of West Virginia, who was originally elected to the Senate in 1958 and served until 2010, as one of the masters of pork-barrel politics, directing a steady stream of federal funds to his home state. A journalist once compiled a list of structures in West Virginia at least partly government funded and named after Byrd: “the Robert C. Byrd Highway; the Robert C. Byrd Locks and Dam; the Robert C. Byrd Institute; the Robert C. Byrd Life Long Learning Center; the Robert C. Byrd Honors Scholarship Program; the Robert C. Byrd Green Bank Telescope; the Robert C. Byrd Institute for Advanced Flexible Manufacturing; the Robert C. Byrd Federal Courthouse; the Robert C. Byrd Health Sciences Center; the Robert C. Byrd Academic and Technology Center; the Robert C. Byrd United Technical Center; the Robert C. Byrd Federal Building; the Robert C. Byrd Drive; the Robert C. Byrd Hilltop Office Complex; the Robert C. Byrd Library; and the Robert C. Byrd Learning Resource Center; the Robert C. Byrd Rural Health Center.” This list does not include government-funded projects in West Virginia that were not named after Byrd. Of course, we would have to analyze each of these expenditures in detail to figure out whether we should treat them as pork-barrel spending or whether they provide widespread benefits that reach beyond West Virginia. At least some of them, or a portion of them, certainly would fall into that category. Because there are currently no term limits for Congressional representatives, those who have been in office longer generally have more power to enact pork-barrel projects.

The amount that government spends on individual pork-barrel projects is small, but many small projects can add up to a substantial total. A nonprofit watchdog organization, called Citizens against Government Waste, produces an annual report, the *Pig Book* that attempts to quantify the amount of pork-barrel spending, focusing on items that only one member of Congress requested, that were passed into law without any public hearings, or that serve only a local purpose. Whether any specific item qualifies as pork can be controversial. The 2021 Congressional Pig Book identified 285 earmarks in FY 2021, with a cost of \$16.8 billion. Recent growth in earmarks and their cost is apparent: in FY 2017, there were 163 earmarks at a cost of \$6.8 billion. Hence, in only four years, there was a 75% increase in the number of earmarks and a 147% increase in the cost of those earmarks.

Logrolling, an action in which all members of a group of legislators agree to vote for a package of otherwise unrelated laws that they individually favor, can encourage pork barrel spending. For example, if one member of the U.S. Congress suggests building a new bridge or hospital in their own congressional district, the other members might oppose it. However, if 51% of the legislators come together, they can pass a bill that includes a bridge or hospital for every one of their districts.

As a reflection of this interest of legislators in their own districts, the U.S. government has typically spread out its spending on military bases and weapons programs to congressional districts all across the country. In part, the government does this to help create a situation that encourages members of Congress to vote in support of defense spending.

18.3 Flaws in the Democratic System of Government

LEARNING OBJECTIVES

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

- Assess the median voter theory
- Explain the voting cycle
- Analyze the interrelationship between markets and government

Most developed countries today have a democratic system of government: citizens express their opinions through votes and those votes affect the direction of the country. The advantage of democracy over other systems is that it allows everyone in a society an equal say and therefore may reduce the possibility of a small group of wealthy oligarchs oppressing the masses. There is no such thing as a perfect system, and democracy, for all its popularity, is not without its problems, a few of which we will examine here.

We sometimes sum up and oversimplify democracy in two words: “Majority rule.” When voters face three or more choices, however, then voting may not always be a useful way of determining what the majority prefers.

As one example, consider an election in a state where 60% of the population is liberal and 40% is conservative. If there are only two candidates, one from each side, and if liberals and conservatives vote in the same 60–40 proportions in which they are represented in the population, then the liberal will win. What if the election ends up including two liberal candidates and one conservative? It is possible that the liberal vote will split and victory will go to the minority party. In this case, the outcome does not reflect the majority’s preference.

Does the majority view prevail in the case of sugar quotas? Clearly there are more sugar consumers in the United States than sugar producers, but the U.S. domestic sugar lobby (www.sugarcane.org) has successfully argued for protection against imports since 1789. By law, therefore, U.S. cookie and candy makers must use 85% domestic sugar in their products. Meanwhile quotas on imported sugar restrict supply and keep the domestic sugar price up—raising prices for companies that use sugar in producing their goods and for consumers. The European Union allows sugar imports, and prices there are 40% lower than U.S. sugar prices. Sugar-producing countries in the Caribbean repeatedly protest the U.S. quotas at the World Trade Organization meetings, but each bite of cookie, at present, costs you more than if there were no sugar lobby. This case goes against the theory of the “median” voter in a democracy. The **median voter theory** argues that politicians will try to match policies to what pleases the median voter preferences. If we think of political positions along a spectrum from left to right, the median voter is in the middle of the spectrum. This theory argues that actual policy will reflect “middle of the road.” In the case of sugar lobby politics, the *minority*, not the median, dominates policy.

Sometimes it is not even clear how to define the majority opinion. Step aside from politics for a moment and think about a choice facing three families (the Ortegas, the Schmidts, and the Alexanders) who are planning to celebrate New Year’s Day together. They agree to vote on the menu, choosing from three entrees, and they agree that the majority vote wins. With three families, it seems reasonable that one producing choice will get a 2–1 majority. What if, however, their vote ends up looking like [Table 18.1](#)?

Clearly, the three families disagree on their first choice. However, the problem goes even deeper. Instead of looking at all three choices at once, compare them two at a time. (See [Figure 18.2](#)) In a vote of turkey versus beef, turkey wins by 2–1. In a vote of beef versus lasagna, beef wins 2–1. If turkey beats beef, and beef beats lasagna, then it might seem only logical that turkey must also beat lasagna. However, with the preferences, lasagna is preferred to turkey by a 2–1 vote, as well. If lasagna is preferred to turkey, and turkey beats beef, then surely it must be that lasagna also beats beef? Actually, no. Beef beats lasagna. In other words, the majority view may not win. Clearly, as any car salesperson will tell you, the way one presents choices to us influences our decisions.

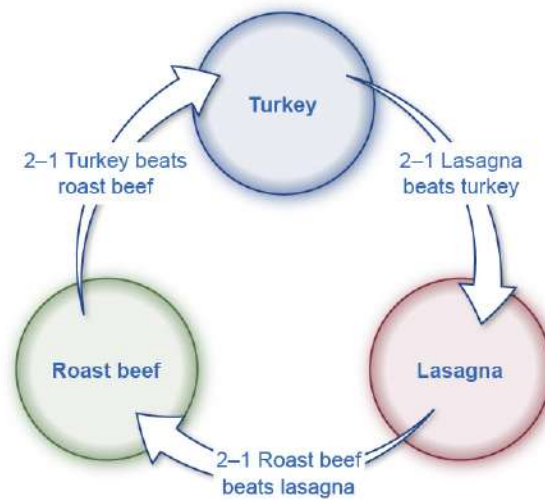


FIGURE 18.2 A Voting Cycle Given these choices, voting will struggle to produce a majority outcome. Turkey is favored over roast beef by 2–1 and roast beef is favored over lasagna by 2–1. If turkey beats roast beef and roast beef beats lasagna, then it might seem that turkey must beat lasagna, too. However, given these preferences, lasagna is favored over turkey by 2–1.

	The Ortega Family	The Schmidt Family	The Alexander Family
First Choice	Turkey	Roast beef	Lasagna
Second Choice	Roast beef	Lasagna	Turkey
Third Choice	Lasagna	Turkey	Roast beef

TABLE 18.1 Circular Preferences

We call the situation in which Choice A is preferred by a majority over Choice B, Choice B is preferred by a majority over Choice C, and Choice C is preferred by a majority over Choice A a **voting cycle**. It is easy to imagine sets of government choices—say, perhaps the choice between increased defense spending, increased government spending on health care, and a tax cut—in which a voting cycle could occur. The result will be determined by the order in which interested parties present and vote on choices, not by majority rule, because every choice is both preferred to some alternative and also not preferred to another alternative.

LINK IT UP

Visit this [website \(http://www.fairvote.org/rcv#rcvbenefits\)](http://www.fairvote.org/rcv#rcvbenefits) to read about ranked choice voting, a preferential voting system.

Where Is Government's Self-Correcting Mechanism?

When a firm produces a product no one wants to buy or produces at a higher cost than its competitors, the firm is likely to suffer losses. If it cannot change its ways, it will go out of business. This self-correcting mechanism in the marketplace can have harsh effects on workers or on local economies, but it also puts pressure on firms for good performance.

Government agencies, however, do not sell their products in a market. They receive tax dollars instead. They are not challenged by competitors as are private-sector firms. If the U.S. Department of Education or the U.S. Department of Defense is performing poorly, citizens cannot purchase their services from another provider and drive the existing government agencies into bankruptcy. If you are upset that the Internal Revenue Service

is slow in sending you a tax refund or seems unable to answer your questions, you cannot decide to pay your income taxes through a different organization. Of course, elected politicians can assign new leaders to government agencies and instruct them to reorganize or to emphasize a different mission. The pressure government faces, however, to change its bureaucracy, to seek greater efficiency, and to improve customer responsiveness is much milder than the threat of being put out of business altogether.

This insight suggests that when government provides goods or services directly, we might expect it to do so with less efficiency than private firms—except in certain cases where the government agency may compete directly with private firms. At the local level, for example, government can provide directly services like garbage collection, using private firms under contract to the government, or by a mix of government employees competing with private firms.

A Balanced View of Markets and Government

The British statesman Sir Winston Churchill (1874–1965) once wrote: “No one pretends that democracy is perfect or all-wise. Indeed, it has been said that democracy is the worst form of government except for all of the other forms which have been tried from time to time.” In that spirit, the theme of this discussion is certainly not that we should abandon democratic government. A practical student of public policy needs to recognize that in some cases, like the case of well-organized special interests or pork-barrel legislation, a democratic government may seek to enact economically unwise projects or programs. In other cases, by placing a low priority on the problems of those who are not well organized or who are less likely to vote, the government may fail to act when it could do some good. In these and other cases, there is no automatic reason to believe that government will necessarily make economically sensible choices.

“The true test of a first-rate mind is the ability to hold two contradictory ideas at the same time,” wrote the American author F. Scott Fitzgerald (1896–1940). At this point in your study of microeconomics, you should be able to go one better than Fitzgerald and hold three somewhat contradictory ideas about the interrelationship between markets and government in your mind at the same time.

First, markets are extraordinarily useful and flexible institutions through which society can allocate its scarce resources. We introduced this idea with the subjects of international trade and demand and supply in other chapters and reinforced it in all the subsequent discussions of how households and firms make decisions.

Second, markets may sometimes produce unwanted results. A short list of the cases in which markets produce unwanted results includes monopoly and other cases of imperfect competition, pollution, poverty and inequality of incomes, discrimination, and failure to provide insurance.

Third, while government may play a useful role in addressing the problems of markets, government action is also imperfect and may not reflect majority views. Economists readily admit that, in settings like monopoly or negative externalities, a potential role exists for government intervention. However, in the real world, it is not enough to point out that government action might be a good idea. Instead, we must have some confidence that the government is likely to identify and carry out the appropriate public policy. To make sensible judgments about economic policy, we must see the strengths and weaknesses of both markets and government. We must not idealize or demonize either unregulated markets or government actions. Instead, consider the actual strengths and weaknesses of real-world markets and real-world governments.

These three insights seldom lead to simple or obvious political conclusions. As the famous British economist Joan Robinson wrote some decades ago: “[E]conomic theory, in itself, preaches no doctrines and cannot establish any universally valid laws. It is a method of ordering ideas and formulating questions.” The study of economics is neither politically conservative, nor moderate, nor liberal. There are economists who are Democrats, Republicans, libertarians, socialists, and members of every other political group you can name. Of course, conservatives may tend to emphasize the virtues of markets and the limitations of government, while liberals may tend to emphasize the shortcomings of markets and the need for government programs. Such differences only illustrate that the language and terminology of economics is not limited to one set of political

beliefs, but can be used by all.



BRING IT HOME

Chinese Tire Tariffs

In April 2009, the union representing U.S. tire manufacturing workers filed a request with the U.S. International Trade Commission (ITC), asking it to investigate tire imports from China. Under U.S. trade law, if imports from a country increase to the point that they cause market disruption in the United States, as determined by the ITC, then it can also recommend a remedy for this market disruption. In this case, the ITC determined that from 2004 to 2008, U.S. tire manufacturers suffered declines in production, financial health, and employment as a direct result of increases in tire imports from China. The ITC recommended placing an additional tax on tire imports from China. President Obama and Congress agreed with the ITC recommendation, and in June 2009 tariffs on Chinese tires increased from 4% to 39%. In addition, tariffs on Chinese tires increased further as part of President Trump's increases on a broad range of Chinese products.

Why would U.S. consumers buy imported tires from China in the first place? Most likely, because they are cheaper than tires produced domestically or in other countries. Therefore, this tariff increase should cause U.S. consumers to pay higher prices for tires, either because Chinese tires are now more expensive, or because U.S. consumers are pushed by the tariff to buy more expensive tires made by U.S. manufacturers or those from other countries. In the end, this tariff made U.S. consumers pay more for tires.

Was this tariff met with outrage expressed via social media, traditional media, or mass protests? Were there “Occupy Wall Street-type” demonstrations? The answer is a resounding “No”. Most U.S. tire consumers were likely unaware of the tariff increase, although they may have noticed the price increase, which was between \$4 and \$13 depending on the type of tire. Tire consumers are also potential voters. Conceivably, a tax increase, even a small one, might make voters unhappy. However, voters probably realized that it was not worth their time to learn anything about this issue or cast a vote based on it. They probably thought their vote would not matter in determining the outcome of an election or changing this policy.

Estimates of the impact of this tariff show it costs U.S. consumers around \$1.11 billion annually. Of this amount, roughly \$817 million ends up in the pockets of foreign tire manufacturers other than in China, and the remaining \$294 million goes to U.S. tire manufacturers. In other words, the tariff increase on Chinese tires may have saved 1,200 jobs in the domestic tire sector, but it cost 3,700 jobs in other sectors, as consumers had to reduce their spending because they were paying more for tires. People actually lost their jobs as a result of this tariff. Workers in U.S. tire manufacturing firms earned about \$40,000 in 2010. Given the number of jobs saved and the total cost to U.S. consumers, the cost of saving one job amounted to \$926,500!

This tariff caused a net decline in U.S. social surplus. (We discuss total surplus in the [Demand and Supply](#) chapter, and tariffs in the [Introduction to International Trade](#) chapter.) Instead of saving jobs, it cost jobs, and those jobs that it saved cost many times more than the people working in them could ever hope to earn. Why would the government do this?

The chapter answers this question by discussing the influence special interest groups have on economic policy. The steelworkers union, whose members make tires, saw increasingly more members lose their jobs as U.S. consumers consumed increasingly more cheap Chinese tires. By definition, this union is relatively small but well organized, especially compared to tire consumers. It stands to gain much for each of its members, compared to what each tire consumer may have to give up in terms of higher prices. Thus, the steelworkers union (joined by domestic tire manufacturers) has not only the means but the incentive to lobby economic policymakers and lawmakers. Given that U.S. tire consumers are a large and unorganized group, if they even are a group, it is unlikely they will lobby against higher tire tariffs. In the end, lawmakers tend to listen to those who lobby them, even though the results make for bad economic policy.

Key Terms

logrolling the situation in which groups of legislators all agree to vote for a package of otherwise unrelated laws that they individually favor

median voter theory theory that politicians will try to match policies to what pleases the median voter preferences

pork-barrel spending spending that benefits mainly a single political district

rational ignorance the theory that rational people will not vote if the costs of becoming informed and voting are too high or because they know their vote will not be decisive in the election

special interest groups groups that are small in number relative to the nation, but well organized and thus exert a disproportionate effect on political outcomes

voting cycle the situation in which a majority prefers A over B, B over C, and C over A

Key Concepts and Summary

18.1 Voter Participation and Costs of Elections

The theory of rational ignorance says voters will recognize that their single vote is extremely unlikely to influence the outcome of an election. As a consequence, they will choose to remain uninformed about issues and not vote. This theory helps explain why voter turnout is so low in the United States.

18.2 Special Interest Politics

Special interest politics arises when a relatively small group, called a special interest group, each of whose members has a large interest in a political outcome, devotes considerable time and energy to lobbying for the group's preferred choice. Meanwhile, the large majority, each of whose members has only a small interest in this issue, pays no attention.

We define pork--barrel spending as legislation whose benefits are concentrated on a single district while the costs are spread widely over the country. Logrolling refers to a situation in which two or more legislators agree to vote for each other's legislation, which can then encourage pork-barrel spending in many districts.

18.3 Flaws in the Democratic System of Government

Majority votes can run into difficulties when more than two choices exist. A voting cycle occurs when, in a situation with at least three choices, choice A is preferred by a majority vote to choice B, choice B is preferred by a majority vote to choice C, and choice C is preferred by a majority vote to choice A. In such a situation, it is impossible to identify what the majority prefers. Another difficulty arises when the vote is so divided that no choice receives a majority.

A practical approach to microeconomic policy will need to take a realistic view of the specific strengths and weaknesses of markets as well as government, rather than making the easy but wrong assumption that either the market or government is always beneficial or always harmful.

Self-Check Questions

1. Based on the theory of rational ignorance, what should we expect to happen to voter turnout as the internet makes information easier to obtain?
2. What is the cost of voting in an election?
3. What is the main factor preventing a large community from influencing policy in the same way as a special interest group?
4. Why might legislators vote to impose a tariff on Egyptian cotton, when consumers in their districts would benefit from its availability?

5. True or false: Majority rule can fail to produce a single preferred outcome when there are more than two choices.
6. Anastasia, Emma, and Greta are deciding what to do on a weekend getaway. They each suggest a first, second, and third choice and then vote on the options. [Table 18.2](#) shows their first, second, and third choice preferences. Explain why they will have a hard time reaching a decision. Does the group prefer mountain biking to canoeing? What about canoeing compared to the beach? What about the beach compared to the original choice of mountain biking?

	Anastasia	Emma	Greta
First Choice	Beach	Mountain biking	Canoeing
Second Choice	Mountain biking	Canoeing	Beach
Third Choice	Canoeing	Beach	Mountain biking

TABLE 18.2

7. Suppose there is an election for Soft Drink Commissioner. The field consists of one candidate from the Pepsi party and four from the Coca-Cola party. This would seem to indicate a strong preference for Coca-Cola among the voting population, but the Pepsi candidate ends up winning in a landslide. Why does this happen?

Review Questions

8. How does rational ignorance discourage voting?
9. How can a small special interest group win in a situation of majority voting when the benefits it seeks flow only to a small group?
10. How can pork-barrel spending occur in a situation of majority voting when it benefits only a small group?
11. Why do legislators vote for spending projects in districts that are not their own?
12. Why does a voting cycle make it impossible to decide on a majority-approved choice?
13. How does a government agency raise revenue differently from a private company, and how does that affect the way government makes decisions compared to business decisions?

Critical Thinking Questions

14. What are some reasons people might find acquiring information about politics and voting rational, in contrast to rational ignorance theory?
15. What are some possible ways to encourage voter participation and overcome rational ignorance?
16. Given that rational ignorance discourages some people from becoming informed about elections, is it necessarily a good idea to encourage greater voter turnout? Why or why not?
17. When Microsoft was founded, the company devoted very few resources to lobbying activities. After a high-profile antitrust case against it, however, the company began to lobby heavily. Why does it make financial sense for companies to invest in lobbyists?
18. Representatives of competing firms often comprise special interest groups. Why are competitors sometimes willing to cooperate in order to form lobbying associations?

19. Special interests do not oppose regulations in all cases. The Marketplace Fairness Act of 2013 would require online merchants to collect sales taxes from their customers in other states. Why might a large online retailer like Amazon.com support such a measure?
20. To ensure safety and efficacy, the Food and Drug Administration regulates the medicines that pharmacies are allowed to sell in the United States. Sometimes this means a company must test a drug for years before it can reach the market. We can easily identify the winners in this system as those who are protected from unsafe drugs that might otherwise harm them. Who are the more anonymous losers who do not benefit from strict medical regulations?
21. How is it possible to bear a cost without realizing it? What are some examples of policies that affect people in ways of which they may not even be aware?
22. Is pork-barrel spending always a bad thing? Can you think of some examples of pork-barrel projects, perhaps from your own district, that have had positive results?
23. The United States currently uses a voting system called “first past the post” in elections, meaning that the candidate with the most votes wins. What are some of the problems with a “first past the post” system?
24. What are some alternatives to a “first past the post” system that might reduce the problem of voting cycles?
25. AT&T spent some \$10 million dollars lobbying Congress to block entry of competitors into the telephone market in 1978. Why do you think its efforts failed?
26. Occupy Wall Street was a national (and later global) organized protest against the greed, bank profits, and financial corruption that led to the 2008–2009 recession. The group popularized slogans like “We are the 99%,” meaning it represented the majority against the wealth of the top 1%. Does the fact that the protests had little to no effect on legislative changes support or contradict the chapter?

Problems

27. Say that the government is considering a ban on smoking in restaurants in Tobaccoville. There are 1 million people living there, and each would benefit by \$200 from this smoking ban. However, there are two large tobacco companies in Tobaccoville and the ban would cost them \$5 million each. What are the proposed policy's total costs and benefits? Do you think it will pass?



FIGURE 19.1 The Great Depression At times, such as when many people having trouble making ends meet, it is easy to tell how the economy is doing. This photograph shows people lined up during the Great Depression, waiting for relief checks. At other times, when some are doing well and others are not, it is more difficult to ascertain how the economy of a country is doing. (Credit: modification of “Waiting for relief checks. Calipatria, California” by Dorothea Lange/Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA, Public Domain)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Measuring the Size of the Economy: Gross Domestic Product
- Adjusting Nominal Values to Real Values
- Tracking Real GDP over Time
- Comparing GDP among Countries
- How Well GDP Measures the Well-Being of Society

Introduction to the Macroeconomic Perspective



BRING IT HOME

How is the Economy Doing? How Does One Tell?

The 1990s were boom years for the U.S. economy. Beginning in the late 2000s, from 2007 to 2014, economic performance in the U.S. was poor. The economy experienced another period of strong growth between 2014 and 2019, before COVID-19 rocked the world economy in March and April of 2020. What causes the economy to expand or contract? Why do businesses fail when they are making all the right decisions? Why do workers lose their jobs when they are hardworking and productive? Are bad economic times a failure of the market system? Are they a

failure of the government? These are all questions of macroeconomics, which we will begin to address in this chapter. We will not be able to answer all of these questions here, but we will start with the basics: How is the economy doing? How can we tell?

The macro economy includes all buying and selling, all production and consumption; everything that goes on in every market in the economy. How can we get a handle on that? The answer begins more than 80 years ago, during the Great Depression. President Franklin D. Roosevelt and his economic advisers knew things were bad—but how could they express and measure just how bad it was? An economist named Simon Kuznets, who later won the Nobel Prize for his work, came up with a way to track what the entire economy is producing. In this chapter, you will learn how the government constructs GDP, how we use it, and why it is so important.

Macroeconomics focuses on the economy as a whole (or on whole economies as they interact). What causes recessions? What makes unemployment stay high when recessions are supposed to be over? Why do some countries grow faster than others? Why do some countries have higher standards of living than others? These are all questions that macroeconomics addresses. Macroeconomics involves adding up the economic activity of all households and all businesses in all markets to obtain the overall demand and supply in the economy. However, when we do that, something curious happens. It is not unusual that what results at the macro level is different from the sum of the microeconomic parts. What seems sensible from a microeconomic point of view can have unexpected or counterproductive results at the macroeconomic level. Imagine that you are sitting at an event with a large audience, like a live concert or a basketball game. A few people decide that they want a better view, and so they stand up. However, when these people stand up, they block the view for other people, and the others need to stand up as well if they wish to see. Eventually, nearly everyone is standing up, and as a result, no one can see much better than before. The rational decision of some individuals at the micro level—to stand up for a better view—ended up as self-defeating at the macro level. This is not macroeconomics, but it is an apt analogy.

Macroeconomics is a rather massive subject. How are we going to tackle it? [Figure 19.2](#) illustrates the structure we will use. We will study macroeconomics from three different perspectives:

1. What are the macroeconomic goals? (Macroeconomics as a discipline does not have goals, but we do have goals for the macro economy.)
2. What are the frameworks economists can use to analyze the macroeconomy?
3. Finally, what are the policy tools governments can use to manage the macroeconomy?



FIGURE 19.2 Macroeconomic Goals, Framework, and Policies This chart shows what macroeconomics is about. The box on the left indicates a consensus of what are the most important goals for the macro economy, the middle box lists the frameworks economists use to analyze macroeconomic changes (such as inflation or recession), and the box on the right indicates the two tools the federal government uses to influence the macro economy.

Goals

In thinking about the macroeconomy's overall health, it is useful to consider three primary goals: economic growth, low unemployment, and low inflation.

- Economic growth ultimately determines the prevailing standard of living in a country. Economists measure growth by the percentage change in real (inflation-adjusted) gross domestic product. A growth

rate of more than 3% is considered good.

- Unemployment, as measured by the unemployment rate, is the percentage of people in the labor force who do not have a job. When people lack jobs, the economy is wasting a precious resource—labor, and the result is lower goods and services produced. Unemployment, however, is more than a statistic—it represents people's livelihoods. While measured unemployment is unlikely to ever be zero, economists consider a measured unemployment rate of 5% or less low (good).
- Inflation is a sustained increase in the overall level of prices, and is measured by the consumer price index. If many people face a situation where the prices that they pay for food, shelter, and healthcare are rising much faster than the wages they receive for their labor, there will be widespread unhappiness as their standard of living declines. For that reason, low inflation—an inflation rate of 1–2%—is a major goal.

Frameworks

As you learn in the micro part of this book, principal tools that economists use are theories and models (see [Welcome to Economics!](#) for more on this). In microeconomics, we used the theories of supply and demand. In macroeconomics, we use the theories of aggregate demand (AD) and aggregate supply (AS). This book presents two perspectives on macroeconomics: the Neoclassical perspective and the Keynesian perspective, each of which has its own version of AD and AS. Between the two perspectives, you will obtain a good understanding of what drives the macroeconomy.

Policy Tools

National governments have two tools for influencing the macroeconomy. The first is monetary policy, which involves managing the money supply and interest rates. The second is fiscal policy, which involves changes in government spending/purchases and taxes.

We will explain each of the items in [Figure 19.2](#) in detail in one or more other chapters. As you learn these things, you will discover that the goals and the policy tools are in the news almost every day.

19.1 Measuring the Size of the Economy: Gross Domestic Product

LEARNING OBJECTIVES

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

- Identify the components of GDP on the demand side and on the supply side
- Evaluate how economists measure gross domestic product (GDP)
- Contrast and calculate GDP, net exports, and net national product

Macroeconomics is an empirical subject, so the first step toward understanding it is to measure the economy.

How large is the U.S. economy? Economists typically measure the size of a nation's overall economy by its **gross domestic product (GDP)**, which is the value of all final goods and services produced within a country in a given year. Measuring GDP involves counting the production of millions of different goods and services—smart phones, cars, music downloads, computers, steel, bananas, college educations, and all other new goods and services that a country produced in the current year—and summing them into a total dollar value. This task is straightforward: take the quantity of everything produced, multiply it by the price at which each product sold, and add up the total. In 2020, the U.S. GDP totaled \$20.9 trillion, the largest GDP in the world.

Each of the market transactions that enter into GDP must involve both a buyer and a seller. We can measure an economy's GDP either by the total dollar value of what consumers purchase in the economy, or by the total dollar value of what the country produces. There is even a third way, as we will explain later.

GDP Measured by Components of Demand

Who buys all of this production? We can divide this demand into four main parts: consumer spending (consumption), business spending (investment), government spending on goods and services, and spending

on net exports. (See the following Clear It Up feature to understand what we mean by investment.) [Table 19.1](#) shows how these four components added up to the GDP in 2020, [Figure 19.4](#) (a) shows the levels of consumption, investment, and government purchases over time, expressed as a percentage of GDP, while [Figure 19.4](#) (b) shows the levels of exports and imports as a percentage of GDP over time. A few patterns about each of these components are worth noticing. [Table 19.1](#) shows the components of GDP from the demand side.

	Components of GDP on the Demand Side (in trillions of dollars)	Percentage of Total
Consumption	\$14.0	67.2%
Investment	\$3.6	17.4%
Government	\$3.9	18.5%
Exports	\$2.1	10.2%
Imports	−\$2.7	−13.3%
Total GDP	\$20.9	100%

TABLE 19.1 Components of U.S. GDP in 2022: From the Demand Side (Source: http://bea.gov/iTable/index_nipa.cfm, Table 1.1.5)

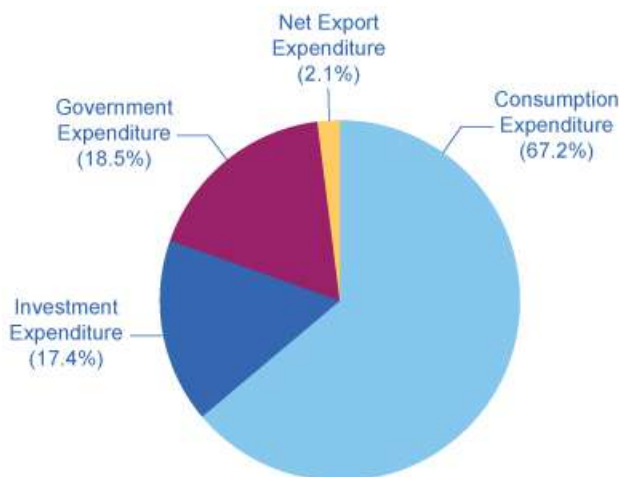


FIGURE 19.3 Percentage of Components of U.S. GDP on the Demand Side Consumption makes up over half of the demand side components of the GDP. Totals in the chart do not add to 100% due to rounding. (Source: http://bea.gov/iTable/index_nipa.cfm, Table 1.1.10)



CLEAR IT UP

What does the word “investment” mean?

What do economists mean by investment, or business spending? In calculating GDP, investment does not refer to purchasing stocks and bonds or trading financial assets. It refers to purchasing new capital goods, that is, new commercial real estate (such as buildings, factories, and stores) and equipment, residential housing construction, and inventories. Inventories that manufacturers produce this year are included in this year’s GDP—even if they are not yet sold. From the accountant’s perspective, it is as if the firm invested in its own inventories. Business investment in 2020 was \$3.6 trillion, according to the Bureau of Economic Analysis.

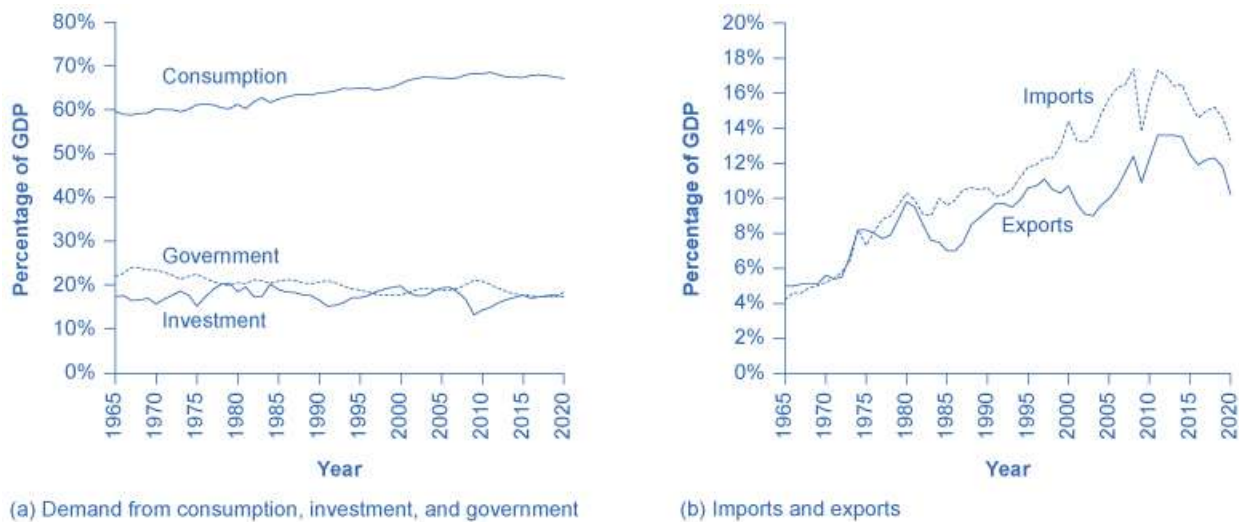


FIGURE 19.4 Components of GDP on the Demand Side (a) Consumption is about two-thirds of GDP, and it has been on a slight upward trend over time. Business investment hovers around 15% of GDP, but it fluctuates more than consumption. Government spending on goods and services is slightly under 20% of GDP and has declined modestly over time. (b) Exports are added to total demand for goods and services, while imports are subtracted from total demand. If exports exceed imports, as in most of the 1960s and 1970s in the U.S. economy, a trade surplus exists. If imports exceed exports, as in recent years, then a trade deficit exists. (Source: http://bea.gov/iTable/index_nipa.cfm, Table 1.1.10)

Consumption expenditure by households is the largest component of GDP, accounting for about two-thirds of the GDP in any year. This tells us that consumers' spending decisions are a major driver of the economy. However, consumer spending is a gentle elephant: when viewed over time, it does not jump around too much, and has increased modestly from about 60% of GDP in the 1960s and 1970s.

Investment expenditure refers to purchases of physical plant and equipment, primarily by businesses. If Starbucks builds a new store, or Amazon buys robots, they count these expenditures under business investment. Investment demand is far smaller than consumption demand, typically accounting for only about 15–18% of GDP, but it is very important for the economy because this is where jobs are created. However, it fluctuates more noticeably than consumption. Business investment is volatile. New technology or a new product can spur business investment, but then confidence can drop and business investment can pull back sharply.

If you have noticed any of the infrastructure projects (new bridges, highways, airports) launched during the 2009 recession, or if you received a stimulus check during the pandemic-induced recession of 2020–2021, you have seen how important government spending can be for the economy. Government expenditure in the United States is close to 20% of GDP, and includes spending by all three levels of government: federal, state, and local. The only part of government spending counted in demand is government purchases of goods or services produced in the economy. Examples include the government buying a new fighter jet for the Air Force (federal government spending), building a new highway (state government spending), or a new school (local government spending). A significant portion of government budgets consists of transfer payments, like unemployment benefits, veteran's benefits, and Social Security payments to retirees. The government excludes these payments from GDP because it does not receive a new good or service in return or exchange. Instead they are transfers of income from taxpayers to others. If you are curious about the awesome undertaking of adding up GDP, read the following Clear It Up feature.



CLEAR IT UP

How do statisticians measure GDP?

Government economists at the Bureau of Economic Analysis (BEA), within the U.S. Department of Commerce, piece together estimates of GDP from a variety of sources.

Once every five years, in the second and seventh year of each decade, the Bureau of the Census carries out a detailed census of businesses throughout the United States. In between, the Census Bureau carries out a monthly survey of retail sales. The government adjusts these figures with foreign trade data to account for exports that are produced in the United States and sold abroad and for imports that are produced abroad and sold here. Once every ten years, the Census Bureau conducts a comprehensive survey of housing and residential finance. Together, these sources provide the main basis for figuring out what is produced for consumers.

For investment, the Census Bureau carries out a monthly survey of construction and an annual survey of expenditures on physical capital equipment.

For what the federal government purchases, the statisticians rely on the U.S. Department of the Treasury. An annual Census of Governments gathers information on state and local governments. Because the government spends a considerable amount at all levels hiring people to provide services, it also tracks a large portion of spending through payroll records that state governments and the Social Security Administration collect.

With regard to foreign trade, the Census Bureau compiles a monthly record of all import and export documents. Additional surveys cover transportation and travel, and make adjustments for financial services that are produced in the United States for foreign customers.

Many other sources contribute to GDP estimates. Information on energy comes from the U.S. Department of Transportation and Department of Energy. The Agency for Health Care Research and Quality collects information on healthcare. Surveys of landlords find out about rental income. The Department of Agriculture collects statistics on farming.

All these bits and pieces of information arrive in different forms, at different time intervals. The BEA melds them together to produce GDP estimates on a quarterly basis (every three months). The BEA then "annualizes" these numbers by multiplying by four. As more information comes in, the BEA updates and revises these estimates. BEA releases the GDP "advance" estimate for a certain quarter one month after a quarter. The "preliminary" estimate comes out one month after that. The BEA publishes the "final" estimate one month later, but it is not actually final. In July, the BEA releases roughly updated estimates for the previous calendar year. Then, once every five years, after it has processed all the results of the latest detailed five-year business census, the BEA revises all of the past GDP estimates according to the newest methods and data, going all the way back to 1929.



LINK IT UP

Visit this [website \(http://openstax.org/l/beafaq\)](http://openstax.org/l/beafaq) to read FAQs on the BEA site. You can even email your own questions!

When thinking about the demand for domestically produced goods in a global economy, it is important to count spending on exports—domestically produced goods that a country sells abroad. Similarly, we must also subtract spending on imports—goods that a country produces in other countries that residents of this country purchase. The GDP net export component is equal to the dollar value of exports (X) minus the dollar value of imports (M), $(X - M)$. We call the gap between exports and imports the **trade balance**. If a country's exports are larger than its imports, then a country has a **trade surplus**. In the United States, exports typically exceeded imports in the 1960s and 1970s, as [Figure 19.4\(b\)](#) shows.

Since the early 1980s, imports have typically exceeded exports, and so the United States has experienced a **trade deficit** in most years. The trade deficit grew quite large in the late 1990s and in the mid-2000s. [Figure 19.4](#) (b) also shows that imports and exports have both risen substantially in recent decades, even after the declines during the Great Recession between 2008 and 2009. As we noted before, if exports and imports are equal, foreign trade has no effect on total GDP. However, even if exports and imports are balanced overall, foreign trade might still have powerful effects on particular industries and workers by causing nations to shift workers and physical capital investment toward one industry rather than another.

Based on these four components of demand, we can measure GDP as:

$$\text{GDP} = \text{Consumption} + \text{Investment} + \text{Government} + \text{Trade balance}$$

$$\text{GDP} = C + I + G + (X - M)$$

Understanding how to measure GDP is important for analyzing connections in the macro economy and for thinking about macroeconomic policy tools.

GDP Measured by What is Produced

Everything that we purchase somebody must first produce. [Table 19.2](#) breaks down what a country produces into five categories: **durable goods**, **nondurable goods**, **services**, **structures**, and the change in **inventories**. Before going into detail about these categories, notice that total GDP measured according to what is produced is exactly the same as the GDP measured by looking at the five components of demand. [Figure 19.5](#) provides a visual representation of this information.

Components of GDP on the Supply Side (in trillions of dollars)		Percentage of Total
Goods		
Durable goods	\$3.5	16.7%
Nondurable goods	\$2.8	13.4%
Services	\$12.7	60.8%
Structures	\$1.9	9.1%
Change in inventories	\$0.0	0.0%
Total GDP	\$20.9	100%

TABLE 19.2 Components of U.S. GDP on the Production Side, 2020 (Source: http://bea.gov/iTable/index_nipa.cfm, Table 1.2.5)

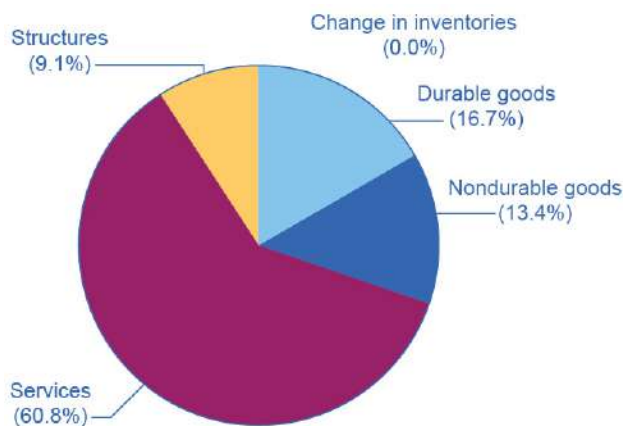


FIGURE 19.5 Percentage of Components of GDP on the Production Side Services make up over 60 percent of the production side components of GDP in the United States.

Since every market transaction must have both a buyer and a seller, GDP must be the same whether measured by what is demanded or by what is produced. [Figure 19.6](#) shows these components of what is produced, expressed as a percentage of GDP, since 1950.

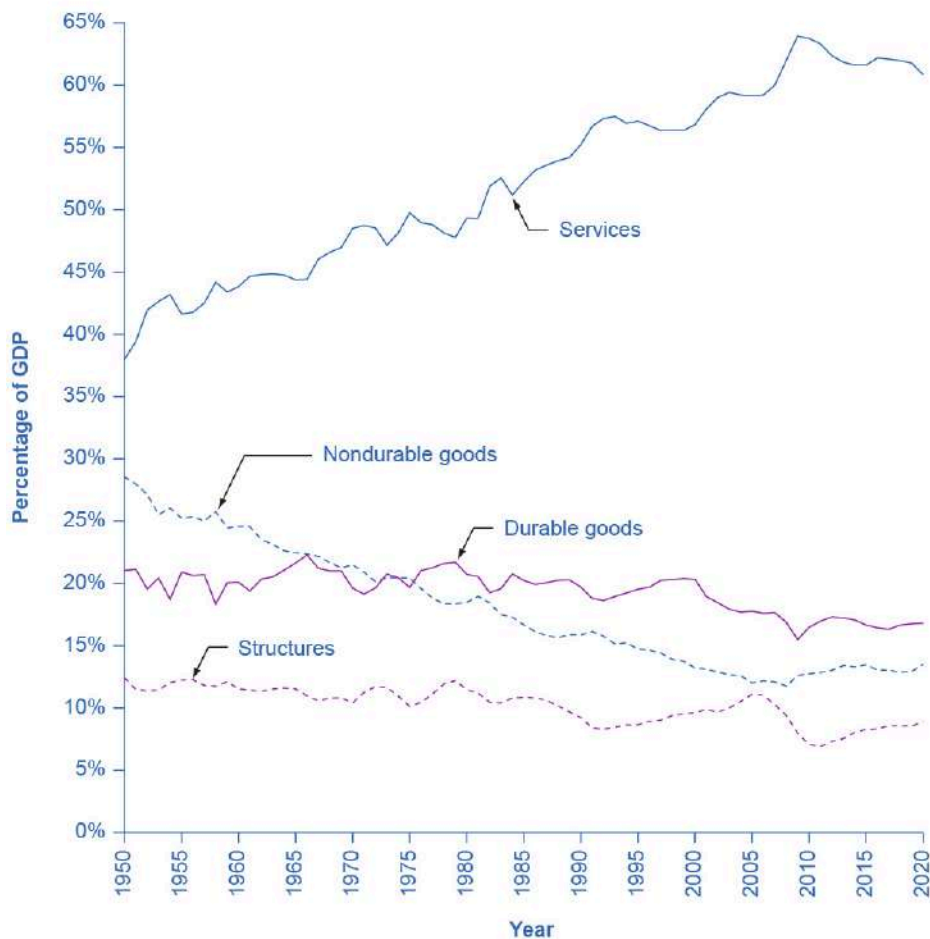


FIGURE 19.6 Types of Production Services are the largest single component of total supply, representing over 60 percent of GDP, up from about 45 percent in the early 1950s. Durable and nondurable goods constitute the manufacturing sector, and they have declined from 40 percent of GDP in 1950 to about 30 percent in 2016. Nondurable goods used to be larger than durable goods, but in recent years, nondurable goods have been dropping

to below the share of durable goods, which is less than 20% of GDP. Structures hover around 10% of GDP. We do not show here the change in inventories, the final component of aggregate supply. It is typically less than 1% of GDP.

In thinking about what is produced in the economy, many non-economists immediately focus on solid, long-lasting goods, like cars and computers. By far the largest part of GDP, however, is services. Moreover, services have been a growing share of GDP over time. A detailed breakdown of the leading service industries would include healthcare, education, and legal and financial services. It has been decades since most of the U.S. economy involved making solid objects. Instead, the most common jobs in a modern economy involve a worker looking at pieces of paper or a computer screen; meeting with co-workers, customers, or suppliers; or making phone calls.

Even within the overall category of goods, long-lasting durable goods like cars and refrigerators are about the same share of the economy as short-lived nondurable goods like food and clothing. The category of structures includes everything from homes, to office buildings, shopping malls, and factories. Inventories is a small category that refers to the goods that one business has produced but has not yet sold to consumers, and are still sitting in warehouses and on shelves. The amount of inventories sitting on shelves tends to decline if business is better than expected, or to rise if business is worse than expected.

Another Way to Measure GDP: The National Income Approach

GDP is a measure of what is produced in a nation. The primary way GDP is estimated is with the Expenditure Approach we discussed above, but there is another way. Everything a firm produces, when sold, becomes revenues to the firm. Businesses use revenues to pay their bills: Wages and salaries for labor, interest and dividends for capital, rent for land, profit to the entrepreneur, etc. So adding up all the income produced in a year provides a second way of measuring GDP. This is why the terms GDP and **national income** are sometimes used interchangeably. The total value of a nation's output is equal to the total value of a nation's income.

The Problem of Double Counting

We define GDP as the current value of all final goods and services produced in a nation in a year. What are final goods? They are goods at the furthest stage of production at the end of a year. Statisticians who calculate GDP must avoid the mistake of **double counting**, in which they count output more than once as it travels through the production stages. For example, imagine what would happen if government statisticians first counted the value of tires that a tire manufacturer produces, and then counted the value of a new truck that an automaker sold that contains those tires. In this example, the statisticians would have counted the value of the tires twice—because the truck's price includes the value of the tires.

To avoid this problem, which would overstate the size of the economy considerably, government statisticians count just the value of **final goods and services** in the chain of production that are sold for consumption, investment, government, and trade purposes. Statisticians exclude **intermediate goods**, which are goods that go into producing other goods, from GDP calculations. From the example above, they will only count the Ford truck's value. The value of what businesses provide to other businesses is captured in the final products at the end of the production chain.

The concept of GDP is fairly straightforward: it is just the dollar value of all final goods and services produced in the economy in a year. In our decentralized, market-oriented economy, actually calculating the more than \$21 trillion-dollar U.S. GDP—along with how it is changing every few months—is a full-time job for a brigade of government statisticians.

What is Counted in GDP	What is not included in GDP
Consumption	Intermediate goods
Business investment	Transfer payments and non-market activities
Government spending on goods and services	Used goods
Net exports	Illegal goods

TABLE 19.3 Counting GDP

Notice the items that are not counted into GDP, as [Table 19.3](#) outlines. The sales of used goods are not included because they were produced in a previous year and are part of that year's GDP. The entire underground economy of services paid “under the table” and illegal sales should be counted, but is not, because it is impossible to track these sales. In Friedrich Schneider's recent study of shadow economies, he estimated the underground economy in the United States to be 6.6% of GDP, or close to \$2 trillion dollars in 2013 alone. Transfer payments, such as payment by the government to individuals, are not included, because they do not represent production. Also, production of some goods—such as home production as when you make your breakfast—is not counted because these goods are not sold in the marketplace.

LINK IT UP

Visit this [website \(http://openstax.org/l/undergroundecon\)](http://openstax.org/l/undergroundecon) to read about the “New Underground Economy.”

Other Ways to Measure the Economy

Besides GDP, there are several different but closely related ways of measuring the size of the economy. We mentioned above that we can think of GDP as total production and as total purchases. We can also think of it as total income since anything one produces and sells yields income.

One of the closest cousins of GDP is the **gross national product (GNP)**. GDP includes only what country produces within its borders. GNP adds what domestic businesses and labor abroad produces, and subtracts any payments that foreign labor and businesses located in the United States send home to other countries. In other words, GNP is based more on what a country's citizens and firms produce, wherever they are located, and GDP is based on what happens within a certain country's geographic boundaries. For the United States, the gap between GDP and GNP is relatively small; in recent years, only about 0.2%. For small nations, which may have a substantial share of their population working abroad and sending money back home, the difference can be substantial.

We calculate **net national product (NNP)** by taking GNP and then subtracting the value of how much physical capital is worn out, or reduced in value because of aging, over the course of a year. The process by which capital ages and loses value is called **depreciation**. We can further subdivide NNP into **national income**, which includes all income to businesses and individuals, and **personal income**, which includes only income to people.

The gross national income (GNI) includes the value of all goods and services produced by people from a country—whether in the country or not. Unlike the other methods, GNI essentially measures the wealth of a nation because it focuses on income, not output. As you will see in the discussion regarding global economic diversity, the World Bank now uses GNI to classify nations according to economic status.

For practical purposes, it is not vital to memorize these definitions. However, it is important to be aware that these differences exist and to know what statistic you are examining, so that you do not accidentally compare, say, GDP in one year or for one country with GNP or NNP in another year or another country. To get an idea of

how these calculations work, follow the steps in the following Work It Out feature.

WORK IT OUT

Calculating GDP, Net Exports, and NNP

Based on the information in [Table 19.4](#):

- What is the value of GDP?
- What is the value of net exports?
- What is the value of NNP?

Government purchases	\$120 billion
Depreciation	\$40 billion
Consumption	\$400 billion
Business Investment	\$60 billion
Exports	\$100 billion
Imports	\$120 billion
Income receipts from rest of the world	\$10 billion
Income payments to rest of the world	\$8 billion

TABLE 19.4

Step 1. To calculate GDP use the following formula:

$$\begin{aligned}
 \text{GDP} &= \text{Consumption} + \text{Investment} + \text{Government spending} + (\text{Exports} - \text{Imports}) \\
 &= C + I + G + (X - M) \\
 &= \$400 + \$60 + \$120 + (\$100 - \$120) \\
 &= \$560 \text{ billion}
 \end{aligned}$$

Step 2. To calculate net exports, subtract imports from exports.

$$\begin{aligned}
 \text{Net exports} &= X - M \\
 &= \$100 - \$120 \\
 &= -\$20 \text{ billion}
 \end{aligned}$$

Step 3. To calculate NNP, use the following formula:

$$\begin{aligned}
 \text{NNP} &= \text{GDP} + \text{Income receipts from the rest of the world} \\
 &\quad - \text{Income payments to the rest of the world} - \text{Depreciation} \\
 &= \$560 + \$10 - \$8 - \$40 \\
 &= \$522 \text{ billion}
 \end{aligned}$$

19.2 Adjusting Nominal Values to Real Values

LEARNING OBJECTIVES

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

- Contrast nominal GDP and real GDP
- Explain GDP deflator
- Calculate real GDP based on nominal GDP values

When examining economic statistics, there is a crucial distinction worth emphasizing. The distinction is between nominal and real measurements, which refer to whether or not inflation has distorted a given statistic. Looking at economic statistics without considering inflation is like looking through a pair of binoculars and trying to guess how close something is: unless you know how strong the lenses are, you cannot guess the distance very accurately. Similarly, if you do not know the inflation rate, it is difficult to figure out if a rise in GDP is due mainly to a rise in the overall level of prices or to a rise in quantities of goods produced. The **nominal value** of any economic statistic means that we measure the statistic in terms of actual prices that exist at the time. The **real value** refers to the same statistic after it has been adjusted for inflation. Generally, it is the real value that is more important.

Converting Nominal to Real GDP

[Table 19.5](#) shows U.S. GDP at five-year intervals since 1960 in nominal dollars; that is, GDP measured using the actual market prices prevailing in each stated year. [Figure 19.7](#) also reflects this data in a graph.

Year	Nominal GDP (billions of dollars)	GDP Deflator (2005 = 100)
1960	542.4	16.6
1965	742.3	17.8
1970	1,073.3	21.7
1975	1,684.9	29.8
1980	2,857.3	42.2
1985	4,339.0	54.5
1990	5,963.1	63.6
1995	7,639.7	71.8
2000	10,251.0	78.0
2005	13,039.2	87.5
2010	15,049.0	96.2
2015	18,206.0	104.7
2020	20,893.7	113.6

TABLE 19.5 U.S. Nominal GDP and the GDP Deflator (Source: <https://apps.bea.gov/itable/index.cfm>, Table 1.1.5 and Table 1.1.9)

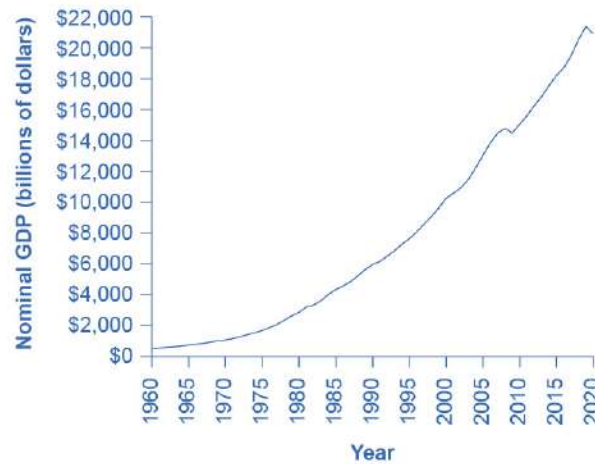


FIGURE 19.7 U.S. Nominal GDP, 1960–2020 Nominal GDP values have risen exponentially from 1960 through 2020, according to the BEA.

If an unwary analyst compared nominal GDP in 1960 to nominal GDP in 2010, it might appear that national output had risen by a factor of more than 38 over this time (that is, GDP of \$20.9 trillion in 2020 divided by GDP of \$543 billion in 1960 = 38). This conclusion would be highly misleading. Recall that we define nominal GDP as the quantity of every final good or service produced multiplied by the price at which it was sold, summed up for all goods and services. In order to see how much production has actually increased, we need to extract the effects of higher prices on nominal GDP. We can easily accomplish this using the GDP deflator.

The GDP deflator is a price index measuring the average prices of all final goods and services included in the economy. We explore price indices in detail and how we compute them in [Inflation](#), but this definition will do in the context of this chapter. [Table 19.5](#) provides the GDP deflator data and [Figure 19.8](#) shows it graphically.

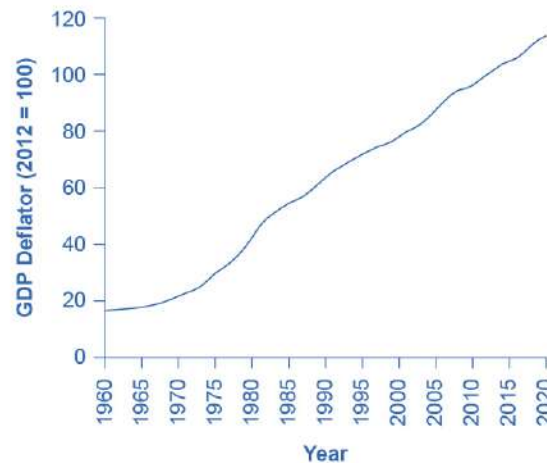


FIGURE 19.8 U.S. GDP Deflator, 1960–2020 Much like nominal GDP, the GDP deflator has risen exponentially from 1960 through 2010. (Source: BEA <https://apps.bea.gov/itable/index.cfm>, Table 1.1.9)

[Figure 19.8](#) shows that the price level has risen dramatically since 1960. The price level in 2020 was seven times higher than in 1960 (the deflator for 2020 was 113 versus a level of 17 in 1960). Clearly, much of the growth in nominal GDP was due to inflation, not an actual change in the quantity of goods and services produced, in other words, not in real GDP. Recall that nominal GDP can rise for two reasons: an increase in output, and/or an increase in prices. What is needed is to extract the increase in prices from nominal GDP so as to measure only changes in output. After all, the dollars used to measure nominal GDP in 1960 are worth more than the inflated dollars of 2020—and the price index tells exactly how much more. This adjustment is

easy to do if you understand that nominal measurements are in value terms, where

$$\text{Value} = \text{Price} \times \text{Quantity}$$

or

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

Let's look at an example at the micro level. Suppose the t-shirt company, Coolshirts, sells 10 t-shirts at a price of \$9 each.

$$\begin{aligned}\text{Coolshirt's nominal revenue from sales} &= \text{Price} \times \text{Quantity} \\ &= \$9 \times 10 \\ &= \$90\end{aligned}$$

Then,

$$\begin{aligned}\text{Coolshirt's real income} &= \frac{\text{Nominal revenue}}{\text{Price}} \\ &= \frac{\$90}{\$9} \\ &= 10\end{aligned}$$

In other words, when we compute “real” measurements we are trying to obtain actual quantities, in this case, 10 t-shirts.

With GDP, it is just a tiny bit more complicated. We start with the same formula as above:

$$\text{Real GDP} = \frac{\text{Nominal GDP}}{\text{Price Index}}$$

For reasons that we will explain in more detail below, mathematically, a price index is a two-digit decimal number like 1.00 or 0.85 or 1.25. Because some people have trouble working with decimals, when the price index is published, it has traditionally been multiplied by 100 to get integer numbers like 100, 85, or 125. What this means is that when we “deflate” nominal figures to get real figures (by dividing the nominal by the price index). We also need to remember to divide the published price index by 100 to make the math work. Thus, the formula becomes:

$$\text{Real GDP} = \frac{\text{Nominal GDP}}{\text{Price Index} / 100}$$

Now read the following Work It Out feature for more practice calculating real GDP.

WORK IT OUT

Computing GDP

It is possible to use the data in [Table 19.5](#) to compute real GDP.

Step 1. Look at [Table 19.5](#), to see that, in 1960, nominal GDP was \$543.3 billion and the price index (GDP deflator) was 19.0.

Step 2. To calculate the real GDP in 1960, use the formula:

$$\begin{aligned}\text{Real GDP} &= \frac{\text{Nominal GDP}}{\text{Price Index} / 100} \\ &= \frac{\$543.3 \text{ billion}}{19 / 100} \\ &= \$2,859.5 \text{ billion}\end{aligned}$$

We'll do this in two parts to make it clear. First adjust the price index: 19 divided by 100 = 0.19. Then divide into nominal GDP: \$543.3 billion / 0.19 = \$2,859.5 billion.

Step 3. Use the same formula to calculate the real GDP in 1965.

$$\begin{aligned}
 \text{Real GDP} &= \frac{\text{Nominal GDP}}{\text{Price Index} / 100} \\
 &= \frac{\$743.7 \text{ billion}}{20.3 / 100} \\
 &= \$3,663.5 \text{ billion}
 \end{aligned}$$

Step 4. Continue using this formula to calculate all of the real GDP values from 1960 through 2010. The calculations and the results are in [Table 19.6](#).

Year	Nominal GDP (billions of dollars)	GDP Deflator (2005 = 100)	Calculations	Real GDP (billions of 2005 dollars)
1960	543.3	19.0	543.3 / (19.0/100)	2859.5
1965	743.7	20.3	743.7 / (20.3/100)	3663.5
1970	1075.9	24.8	1,075.9 / (24.8/100)	4338.3
1975	1688.9	34.1	1,688.9 / (34.1/100)	4952.8
1980	2862.5	48.3	2,862.5 / (48.3/100)	5926.5
1985	4346.7	62.3	4,346.7 / (62.3/100)	6977.0
1990	5979.6	72.7	5,979.6 / (72.7/100)	8225.0
1995	7664.0	82.0	7,664 / (82.0/100)	9346.3
2000	10289.7	89.0	10,289.7 / (89.0/100)	11561.5
2005	13095.4	100.0	13,095.4 / (100.0/100)	13095.4
2010	14958.3	110.0	14,958.3 / (110.0/100)	13598.5

TABLE 19.6 Converting Nominal to Real GDP (Source: Bureau of Economic Analysis, www.bea.gov)

There are a couple things to notice here. Whenever you compute a real statistic, one year (or period) plays a special role. It is called the base year (or base period). The base year is the year whose prices we use to compute the real statistic. When we calculate real GDP, for example, we take the quantities of goods and services produced in each year (for example, 1960 or 1973) and multiply them by their prices in the base year (in this case, 2005), so we get a measure of GDP that uses prices that do not change from year to year. That is why real GDP is labeled “Constant Dollars” or, in this example, “2005 Dollars,” which means that real GDP is constructed

using prices that existed in 2005. While the example here uses 2005 as the base year, more generally, you can use any year as the base year. The formula is:

$$\text{GDP deflator} = \frac{\text{Nominal GDP}}{\text{Real GDP}} \times 100$$

Rearranging the formula and using the data from 2005:

$$\begin{aligned}\text{Real GDP} &= \frac{\text{Nominal GDP}}{\text{Price Index} / 100} \\ &= \frac{\$13,095.4 \text{ billion}}{100 / 100} \\ &= \$13,095.4 \text{ billion}\end{aligned}$$

Comparing real GDP and nominal GDP for 2005, you see they are the same. This is no accident. It is because we have chosen 2005 as the “base year” in this example. Since the price index in the base year always has a value of 100 (by definition), nominal and real GDP are always the same in the base year.

Look at the data for 2010.

$$\begin{aligned}\text{Real GDP} &= \frac{\text{Nominal GDP}}{\text{Price Index} / 100} \\ &= \frac{\$14,958.3 \text{ billion}}{110 / 100} \\ &= \$13,598.5 \text{ billion}\end{aligned}$$

Use this data to make another observation: As long as inflation is positive, meaning prices increase on average from year to year, real GDP should be less than nominal GDP in any year after the base year. The reason for this should be clear: The value of nominal GDP is “inflated” by inflation. Similarly, as long as inflation is positive, real GDP should be greater than nominal GDP in any year before the base year.

Figure 19.9 shows the U.S. nominal and real GDP since 1960. Because 2005 is the base year, the nominal and real values are exactly the same in that year. However, over time, the rise in nominal GDP looks much larger than the rise in real GDP (that is, the nominal GDP line rises more steeply than the real GDP line), because the presence of inflation, especially in the 1970s exaggerates the rise in nominal GDP.

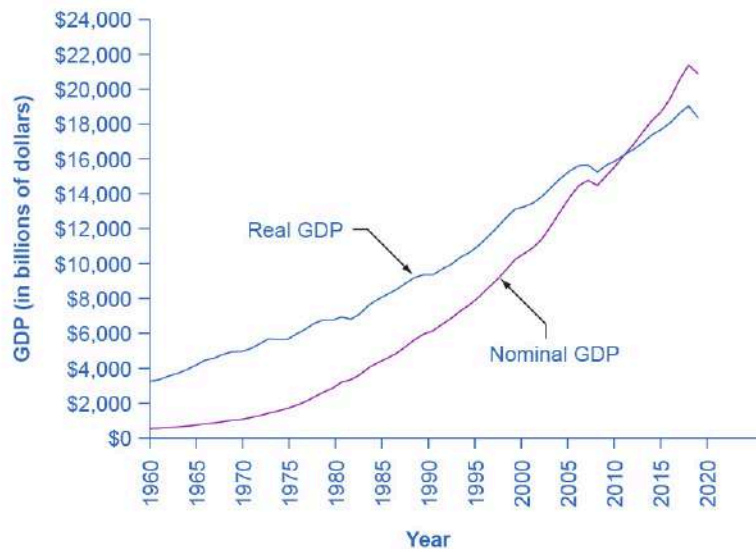


FIGURE 19.9 U.S. Nominal and Real GDP, 1960–2020 The red line measures U.S. GDP in nominal dollars. The black line measures U.S. GDP in real dollars, where all dollar values are converted to 2012 dollars. Since we express real GDP in 2012 dollars, the two lines cross in 2012. However, real GDP will appear higher than nominal GDP in the years before 2012, because dollars were worth less in 2012 than in previous years. Conversely, real GDP will appear

lower in the years after 2012, because dollars were worth more in 2012 than in later years.

Let's return to the question that we posed originally: How much did GDP increase in real terms? What was the real GDP growth rate from 1960 to 2012? To find the real growth rate, we apply the formula for percentage change:

$$\frac{2020 \text{ real GDP} - 1960 \text{ real GDP}}{1960 \text{ real GDP}} \times 100 = \% \text{ change}$$

$$\frac{13,598.5 - 2,859.5}{2,859.5} \times 100 = 376\%$$

In other words, the U.S. economy has increased real production of goods and services by nearly a factor of five since 1960. Of course, that understates the material improvement since it fails to capture improvements in the quality of products and the invention of new products.

There is a quicker way to answer this question approximately, using another math trick. Because:

$$\begin{aligned} \text{Nominal} &= \text{Price} \times \text{Quantity} \\ \% \text{ change in Nominal} &= \% \text{ change in Price} + \% \text{ change in Quantity} \\ \text{OR} \\ \% \text{ change in Quantity} &= \% \text{ change in Nominal} - \% \text{ change in Price} \end{aligned}$$

Therefore, real GDP growth rate (% change in quantity) equals the growth rate in nominal GDP (% change in value) minus the inflation rate (% change in price).

Note that using this equation provides an approximation for small changes in the levels. For more accurate measures, one should use the first formula.

19.3 Tracking Real GDP over Time

LEARNING OBJECTIVES

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

- Explain recessions, depressions, peaks, and troughs
- Evaluate the importance of tracking real GDP over time

When news reports indicate that “the economy grew 1.2% in the first quarter,” the reports are referring to the percentage change in real GDP. By convention, governments report GDP growth at an annualized rate:

Whatever the calculated growth in real GDP was for the quarter, we multiply it by four when it is reported as if the economy were growing at that rate for a full year.

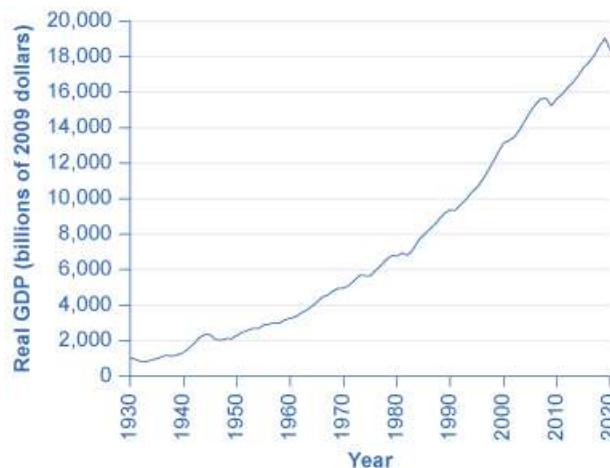


FIGURE 19.10 U.S. GDP, 1930–2020 Real GDP in the United States in 2020 (in 2012 dollars) was about \$18.4 trillion. After adjusting to remove the effects of inflation, this represents a roughly 20-fold increase in the economy's production of goods and services since 1930. (Source: bea.gov)

Figure 19.10 shows the pattern of U.S. real GDP since 1930. Short term declines have regularly interrupted the generally upward long-term path of GDP. We call a significant decline in real GDP a **recession**. We call an especially lengthy and deep recession a **depression**. The severe drop in GDP that occurred during the 1930s Great Depression is clearly visible in the figure, as is the 2008–2009 Great Recession and the recession induced by COVID-19 in 2020.

Real GDP is important because it is highly correlated with other measures of economic activity, like employment and unemployment. When real GDP rises, so does employment.

The most significant human problem associated with recessions (and their larger, uglier cousins, depressions) is that a slowdown in production means that firms need to lay off or fire some of their workers. Losing a job imposes painful financial and personal costs on workers, and often on their extended families as well. In addition, even those who keep their jobs are likely to find that wage raises are scanty at best—or their employers may ask them to take pay cuts.

Table 19.7 lists the pattern of recessions and expansions in the U.S. economy since 1900. We call the highest point of the economy, before the recession begins, the **peak**. Conversely, the lowest point of a recession, before a recovery begins, is the **trough**. Thus, a recession lasts from peak to trough, and an economic upswing runs from trough to peak. We call the economy's movement from peak to trough and trough to peak the **business cycle**. It is intriguing to notice that the three longest trough-to-peak expansions of the twentieth century have happened since 1960. The most recent recession was caused by the COVID-19 pandemic. It started in February 2020 and ended formally in May 2020. This was the most severe recession since the 1930s Great Depression, but also the shortest. The previous recession, called the Great Recession, was also very severe and lasted about 18 months. The expansion starting in June 2009, the trough from the Great Recession, was the longest on record—ending 128 months with the pandemic-induced recession.

Trough	Peak	Months of Contraction	Months of Expansion
December 1900	September 1902	18	21
August 1904	May 1907	23	33
June 1908	January 1910	13	19
January 1912	January 1913	24	12
December 1914	August 1918	23	44
March 1919	January 1920	7	10
July 1921	May 1923	18	22
July 1924	October 1926	14	27
November 1927	August 1929	23	21
March 1933	May 1937	43	50
June 1938	February 1945	13	80
October 1945	November 1948	8	37
October 1949	July 1953	11	45

TABLE 19.7 U.S. Business Cycles since 1900 (Source: <http://www.nber.org/cycles/main.html>)

Trough	Peak	Months of Contraction	Months of Expansion
May 1954	August 1957	10	39
April 1958	April 1960	8	24
February 1961	December 1969	10	106
November 1970	November 1973	11	36
March 1975	January 1980	16	58
July 1980	July 1981	6	12
November 1982	July 1990	16	92
March 1991	March 2001	8	120
November 2001	December 2007	8	73
January 2009	February 2020	2	128
April 2020	TBD	TBD	TBD

TABLE 19.7 U.S. Business Cycles since 1900 (Source: <http://www.nber.org/cycles/main.html>)

A private think tank, the National Bureau of Economic Research (NBER), tracks business cycles for the U.S. economy. However, the effects of a severe recession often linger after the official ending date assigned by the NBER.

19.4 Comparing GDP among Countries

LEARNING OBJECTIVES

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

- Explain how we can use GDP to compare the economic welfare of different nations
- Calculate the conversion of GDP to a common currency by using exchange rates
- Calculate GDP per capita using population data

It is common to use GDP as a measure of economic welfare or standard of living in a nation. When comparing the GDP of different nations for this purpose, two issues immediately arise. First, we measure a country's GDP in its own currency: the United States uses the U.S. dollar; Canada, the Canadian dollar; most countries of Western Europe, the euro; Japan, the yen; Mexico, the peso; and so on. Thus, comparing GDP between two countries requires converting to a common currency. A second issue is that countries have very different numbers of people. For instance, the United States has a much larger economy than Mexico or Canada, but it also has almost three times as many people as Mexico and nine times as many people as Canada. Thus, if we are trying to compare standards of living across countries, we need to divide GDP by population.

Converting Currencies with Exchange Rates

To compare the GDP of countries with different currencies, it is necessary to convert to a “common denominator” using an **exchange rate**, which is the value of one currency in terms of another currency. We express exchange rates either as the units of country A's currency that need to be traded for a single unit of country B's currency (for example, Japanese yen per British pound), or as the inverse (for example, British pounds per Japanese yen). We can use two types of exchange rates for this purpose, market exchange rates

and purchasing power parity (PPP) equivalent exchange rates. Market exchange rates vary on a day-to-day basis depending on supply and demand in foreign exchange markets. PPP-equivalent exchange rates provide a longer run measure of the exchange rate. For this reason, economists typically use PPP-equivalent exchange rates for GDP cross country comparisons. We will discuss exchange rates in more detail in [Exchange Rates and International Capital Flows](#). The following Work It Out feature explains how to convert GDP to a common currency.

WORK IT OUT

Converting GDP to a Common Currency

Using the exchange rate to convert GDP from one currency to another is straightforward. Say that the task is to compare Brazil's GDP in 2020 of 7.4 trillion reals with the U.S. GDP of \$20.9 trillion for the same year.

Step 1. Determine the exchange rate for the specified year. In 2020, the exchange rate was 2.362 reals = \$1. (These numbers are realistic, but rounded off to simplify the calculations.)

Step 2. Convert Brazil's GDP into U.S. dollars:

$$\begin{aligned}\text{Brazil's GDP in \$ U.S.} &= \frac{\text{Brazil's GDP in reals}}{\text{Exchange rate (reals/\$ U.S.)}} \\ &= \frac{7.4 \text{ trillion reals}}{2.362 \text{ reals per \$ U.S.}} \\ &= \$3.1 \text{ trillion}\end{aligned}$$

Step 3. Compare this value to the GDP in the United States in the same year. The U.S. GDP was \$20.9 trillion in 2020, which is almost seven times that of GDP in Brazil.

Step 4. View [Table 19.8](#) which shows the size of and variety of GDPs of different countries in 2020, all expressed in U.S. dollars. We calculate each using the process that we explained above.

Country	GDP in Billions of Domestic Currency		Domestic Currency/U.S. Dollars (PPP Equivalent)	GDP (in billions of U.S. dollars)
Brazil	7,447.86	reals	2.362	3,153.60
Canada	2,204.91	dollars	1.206	1,827.70
China	101,598.62	yuan	4.186	24,273.31
Egypt	5,820.00	pounds	4.511	1,290.21
Germany	3,367.56	euros	0.746	4,516.93
India	195,861.61	rupees	21.990	8,907.02
Japan	531,247.88	yen	102.835	5,166.00
Mexico	23,122.02	pesos	9.522	2,428.20
South Korea	1,924,452.90	won	861.824	2,233.00

TABLE 19.8 Comparing GDPs Across Countries, 2020 (Source: <https://data.worldbank.org/indicator/NY.GDP.MKTP.CN.AD>)

Country	GDP in Billions of Domestic Currency		Domestic Currency/U.S. Dollars (PPP Equivalent)	GDP (in billions of U.S. dollars)
United Kingdom	2,112.04	pounds	0.700	3,019.06
United States	20,936.60	dollars	1.000	20,936.60

TABLE 19.8 Comparing GDPs Across Countries, 2020 (Source: <https://data.worldbank.org/indicator/NY.GDP.MKTP.CN.AD>)

GDP Per Capita

The U.S. economy has the largest GDP in the world, by a considerable amount. The United States is also a populous country; in fact, it is the third largest country by population in the world, although well behind China and India. Is the U.S. economy larger than other countries just because the United States has more people than most other countries, or because the U.S. economy is actually larger on a per-person basis? We can answer this question by calculating a country's **GDP per capita**; that is, the GDP divided by the population.

$$\text{GDP per capita} = \text{GDP/population}$$

The second column of [Table 19.9](#) lists the GDP of the same selection of countries that appeared in the previous [Tracking Real GDP over Time](#) and [Table 19.8](#), showing their GDP as converted into U.S. dollars (which is the same as the last column of the previous table). The third column gives the population for each country. The fourth column lists the GDP per capita. We obtain GDP per capita in two steps: First, by multiplying column two (GDP, in billions of dollars) by 1000 so it has the same units as column three (Population, in millions). Then divide the result (GDP in millions of dollars) by column three (Population, in millions).

Country	GDP (in billions of U.S. dollars)	Population (in millions)	Per Capita GDP (in U.S. dollars)
Brazil	3,153.60	212.56	14,836.27
Canada	1,827.71	38.00	48,097.62
China	24,273.36	1,402.11	17,312.02
Egypt	1,290.21	102.33	12,608.30
Germany	4,516.94	83.24	54,263.99
India	8,907.03	1,380.00	6,454.37
Japan	5,166.00	125.84	41,052.13
Mexico	2,428.20	128.93	18,833.48
South Korea	2,233.00	51.78	43,124.78

TABLE 19.9 GDP Per Capita, 2020 (Source: <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>)

Country	GDP (in billions of U.S. dollars)	Population (in millions)	Per Capita GDP (in U.S. dollars)
United Kingdom	3,019.60	67.22	44,913.08
United States	20,936.60	329.48	63,544.37

TABLE 19.9 GDP Per Capita, 2020 (Source: <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>)

Notice that the rankings by GDP in billions of U.S. dollars, and by GDP per capita, are different than the ranking of GDP by each country's currency. Measured by its own currency, the rupee, India has a somewhat larger GDP than Germany. On a per capita basis in U.S. dollars, Germany has more than 9 times India's per capita GDP on PPP terms.



CLEAR IT UP

Is China going to surpass the United States in terms of standard of living?

China has the largest GDP in PPP terms: \$24 trillion compared to the United States' \$21 trillion. But China has a much larger population so that in per capita terms, its GDP is less than one fourth that of the United States (\$17,000 compared to \$63,000). The Chinese people are still quite poor relative to the United States and other developed countries. One caveat: For reasons we will discuss shortly, GDP per capita can give us only a rough idea of the differences in living standards across countries.

The world's high-income nations—including the United States, Canada, the Western European countries, and Japan—typically have GDP per capita in the range of \$20,000 to \$50,000. Middle-income countries, which include much of Latin America, Eastern Europe, and some countries in East Asia, have GDP per capita in the range of \$6,000 to \$12,000. The world's low-income countries, many of them located in Africa and Asia, often have GDP per capita of less than \$2,000 per year.

19.5 How Well GDP Measures the Well-Being of Society

LEARNING OBJECTIVES

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

- Discuss how productivity influences the standard of living
- Explain the limitations of GDP as a measure of the standard of living
- Analyze the relationship between GDP data and fluctuations in the standard of living

The level of GDP per capita clearly captures some of what we mean by the phrase “standard of living.” Most of the migration in the world, for example, involves people who are moving from countries with relatively low GDP per capita to countries with relatively high GDP per capita.

“Standard of living” is a broader term than GDP. While GDP focuses on production that is bought and sold in markets, **standard of living** includes all elements that affect people's well-being, whether they are bought and sold in the market or not. To illuminate the difference between GDP and standard of living, it is useful to spell out some things that GDP does not cover that are clearly relevant to standard of living.

Limitations of GDP as a Measure of the Standard of Living

GDP measures economic activity, not all activity. As a result, economists like Kate Raworth see it as a somewhat outdated and limited indication of well-being and prosperity. While GDP measures output of work done at home, as well as spending on travel, it doesn't capture unpaid work or leisure time. So, two countries may have equal GDP, but one nation's workers may have an average workday of eight hours, while the other has an average workday of twelve hours. In that case, is their equal GDP truly measuring the prosperity of those

nations? The GDP per capita of the U.S. economy is larger than the GDP per capita of Germany, as [Table 19.9](#) showed, but does that prove that the standard of living in the United States is higher? Not necessarily, since it is also true that the average U.S. worker works several hundred hours more per year more than the average German worker. Calculating GDP does not account for the German worker's extra vacation weeks.

While GDP includes what a country spends on environmental protection, healthcare, and education, it does not include actual levels of environmental cleanliness, health, and learning. GDP includes the cost of buying pollution-control equipment, but it does not address whether the air and water are actually cleaner or dirtier. GDP includes spending on medical care, but does not address whether life expectancy or infant mortality have risen or fallen. Similarly, it counts spending on education, but does not address directly how much of the population can read, write, or do basic mathematics.

GDP includes production that is exchanged in the market, but it does not cover production that is not exchanged in the market. For example, hiring someone to mow your lawn or clean your house is part of GDP, but doing these tasks yourself is not part of GDP. One remarkable change in the U.S. economy in recent decades is the growth in women's participation in the labor force. As of 1970, only about 42% of women participated in the paid labor force. By the second decade of the 2000s, nearly 60% of women participated in the paid labor force according to the Bureau of Labor Statistics. As women are now in the labor force, many of the services they used to produce in the non-market economy like food preparation and child care have shifted to some extent into the market economy, which makes the GDP appear larger even if people actually are not consuming more services. However, as Raworth points out and was explored in the chapter on the labor market, even women who are fully employed expend significant effort (generally more than men) in raising children and maintaining a home. Raworth advocates that economic measures include monetized and un-monetized goods and services, so that the status and contributors to each economy are more accurate.

GDP has nothing to say about the level of inequality in society. GDP per capita is only an average. When GDP per capita rises by 5%, it could mean that GDP for everyone in the society has risen by 5%, or that GDP of some groups has risen by more while that of others has risen by less—or even declined. GDP also has nothing in particular to say about the amount of variety available. If a family buys 100 loaves of bread in a year, GDP does not care whether they are all white bread, or whether the family can choose from wheat, rye, pumpernickel, and many others—it just looks at the total amount the family spends on bread.

Likewise, GDP has nothing much to say about what technology and products are available. The standard of living in, for example, 1950 or 1900 was not affected only by how much money people had—it was also affected by what they could buy. No matter how much money you had in 1950, you could not buy an iPhone or a personal computer.

In certain cases, it is not clear that a rise in GDP is even a good thing. If a city is wrecked by a hurricane, and then experiences a surge of rebuilding construction activity, it would be peculiar to claim that the hurricane was therefore economically beneficial. If people are led by a rising fear of crime, to pay for installing bars and burglar alarms on all their windows, it is hard to believe that this increase in GDP has made them better off. Similarly, some people would argue that sales of certain goods, like pornography or extremely violent movies, do not represent a gain to society's standard of living.

Does a Rise in GDP Overstate or Understate the Rise in the Standard of Living?

The fact that GDP per capita does not fully capture the broader idea of standard of living has led to a concern that the increases in GDP over time are illusory. It is theoretically possible that while GDP is rising, the standard of living could be falling if human health, environmental cleanliness, and other factors that are not included in GDP are worsening. Fortunately, this fear appears to be overstated.

In some ways, the rise in GDP understates the actual rise in the standard of living. For example, the typical workweek for a U.S. worker has fallen over the last century from about 60 hours per week to less than 40 hours per week. Life expectancy and health have risen dramatically, and so has the average level of education. Since

1970, the air and water in the United States have generally been getting cleaner. Companies have developed new technologies for entertainment, travel, information, and health. A much wider variety of basic products like food and clothing is available today than several decades ago. Because GDP does not capture leisure, health, a cleaner environment, the possibilities that new technology creates, or an increase in variety, the actual rise in the standard of living for Americans in recent decades has exceeded the rise in GDP.

On the other side, crime rates, traffic congestion levels, and income inequality are higher in the United States now than they were in the 1960s. Moreover, a substantial number of services that women primarily provided in the non-market economy are now part of the market economy that GDP counts. By ignoring these factors, GDP would tend to overstate the true rise in the standard of living.

LINK IT UP

Visit this [website \(http://openstax.org/l/amdreamvalue\)](http://openstax.org/l/amdreamvalue) to read about the American Dream and standards of living.

GDP is Rough, but Useful

A high level of GDP should not be the only goal of macroeconomic policy, or government policy more broadly. Even though GDP does not measure the broader standard of living with any precision, it does measure production well and it does indicate when a country is materially better or worse off in terms of jobs and incomes. In most countries, a significantly higher GDP per capita occurs hand in hand with other improvements in everyday life along many dimensions, like education, health, and environmental protection.

No single number can capture all the elements of a term as broad as “standard of living.” Nonetheless, GDP per capita is a reasonable, rough-and-ready measure of the standard of living.



BRING IT HOME

How is the Economy Doing? How Does One Tell?

To determine the state of the economy, one needs to examine economic indicators, such as GDP. To calculate GDP is quite an undertaking. It is the broadest measure of a nation’s economic activity and we owe a debt to Simon Kuznets, the creator of the measurement, for that.

The sheer size of the U.S. economy as measured by nominal GDP is huge—as of the third quarter of 2021, \$23.2 trillion worth of goods and services were produced annually. During the COVID-19-induced recession, which lasted just two months according to NBER and was concentrated across Quarters 1 and 2 of 2020, real GDP dropped 9%—much larger and quicker of a drop than during the previous economic downturn, the Great Recession (2007–2009). The economy quickly bounced back, and as of Quarter 1 of 2021, real GDP had slightly surpassed the level it was at prior to the start of the pandemic. These statistics show the severity of the pandemic-induced recession, and while real GDP fully recovered, there are other ways in which the economy has not. While GDP and GDP per capita give us a rough estimate of a nation's standard of living, there are many other ways to track the health of the economy. This chapter is the building block for other chapters that explore more economic indicators such as unemployment, inflation, or interest rates, and perhaps more importantly, will explain how they are related and what causes them to rise or fall.

Key Terms

business cycle the economy's relatively short-term movement in and out of recession

depreciation the process by which capital ages over time and therefore loses its value

depression an especially lengthy and deep decline in output

double counting a potential mistake to avoid in measuring GDP, in which output is counted more than once as it travels through the stages of production

durable good long-lasting good like a car or a refrigerator

exchange rate the price of one currency in terms of another currency

final good and service output used directly for consumption, investment, government, and trade purposes; contrast with “intermediate good”

GDP per capita GDP divided by the population

gross domestic product (GDP) the value of the output of all final goods and services produced within a country in a year

gross national product (GNP) includes what is produced domestically and what is produced by domestic labor and business abroad in a year

intermediate good output provided to other businesses at an intermediate stage of production, not for final users; contrast with “final good and service”

inventory good that has been produced, but not yet been sold

national income includes all income earned: wages, profits, rent, and profit income

net national product (NNP) GNP minus depreciation

nominal value the economic statistic actually announced at that time, not adjusted for inflation; contrast with real value

nondurable good short-lived good like food and clothing

peak during the business cycle, the highest point of output before a recession begins

real value an economic statistic after it has been adjusted for inflation; contrast with nominal value

recession a significant decline in national output

service product which is intangible (in contrast to goods) such as entertainment, healthcare, or education

standard of living all elements that affect people's happiness, whether people buy or sell these elements in the market or not

structure building used as residence, factory, office building, retail store, or for other purposes

trade balance gap between exports and imports

trade deficit exists when a nation's imports exceed its exports and it calculates them as imports – exports

trade surplus exists when a nation's exports exceed its imports and it calculates them as exports – imports

trough during the business cycle, the lowest point of output in a recession, before a recovery begins

Key Concepts and Summary

19.1 Measuring the Size of the Economy: Gross Domestic Product

Economists generally express the size of a nation's economy as its gross domestic product (GDP), which measures the value of the output of all goods and services produced within the country in a year. Economists measure GDP by taking the quantities of all goods and services produced, multiplying them by their prices, and summing the total. Since GDP measures what is bought and sold in the economy, we can measure it either by the sum of what is purchased in the economy or what is produced.

We can divide demand into consumption, investment, government, exports, and imports. We can divide what is produced in the economy into durable goods, nondurable goods, services, structures, and inventories. To avoid double counting, GDP counts only final output of goods and services, not the production of intermediate goods or the value of labor in the chain of production.

19.2 Adjusting Nominal Values to Real Values

The nominal value of an economic statistic is the commonly announced value. The real value is the value after adjusting for changes in inflation. To convert nominal economic data from several different years into real, inflation-adjusted data, the starting point is to choose a base year arbitrarily and then use a price index to convert the measurements so that economists measure them in the money prevailing in the base year.

19.3 Tracking Real GDP over Time

Over the long term, U.S. real GDP have increased dramatically. At the same time, GDP has not increased the same amount each year. The speeding up and slowing down of GDP growth represents the business cycle. When GDP declines significantly, a recession occurs. A longer and deeper decline is a depression. Recessions begin at the business cycle's peak and end at the trough.

19.4 Comparing GDP among Countries

Since we measure GDP in a country's currency, in order to compare different countries' GDPs, we need to convert them to a common currency. One way to do that is with the exchange rate, which is the price of one country's currency in terms of another. Once we express GDPs in a common currency, we can compare each country's GDP per capita by dividing GDP by population. Countries with large populations often have large GDPs, but GDP alone can be a misleading indicator of a nation's wealth. A better measure is GDP per capita.

19.5 How Well GDP Measures the Well-Being of Society

GDP is an indicator of a society's standard of living, but it is only a rough indicator. GDP does not directly take account of leisure, environmental quality, levels of health and education, activities conducted outside the market, changes in inequality of income, increases in variety, increases in technology, or the (positive or negative) value that society may place on certain types of output.

Self-Check Questions

- Country A has export sales of \$20 billion, government purchases of \$1,000 billion, business investment is \$50 billion, imports are \$40 billion, and consumption spending is \$2,000 billion. What is the dollar value of GDP?
- Which of the following are included in GDP, and which are not?
 - The cost of hospital stays
 - The rise in life expectancy over time
 - Child care provided by a licensed day care center
 - Child care provided by a grandmother
 - A used car sale
 - A new car sale
 - The greater variety of cheese available in supermarkets
 - The iron that goes into the steel that goes into a refrigerator bought by a consumer.
- Using data from [Table 19.5](#) how much of the nominal GDP growth from 1980 to 1990 was real GDP and how much was inflation?
- Without looking at [Table 19.7](#), return to [Figure 19.10](#). If we define a recession as a significant decline in national output, can you identify any post-1960 recessions in addition to the 2008-2009 recession? (This requires a judgment call.)
- According to [Table 19.7](#), how often have recessions occurred since the end of World War II (1945)?
- According to [Table 19.7](#), how long has the average recession lasted since the end of World War II?
- According to [Table 19.7](#), how long has the average expansion lasted since the end of World War II?

8. Is it possible for GDP to rise while at the same time per capita GDP is falling? Is it possible for GDP to fall while per capita GDP is rising?
9. The Central African Republic has a GDP of 1,107,689 million CFA francs and a population of 4.862 million. The exchange rate is 284.681 CFA francs per dollar. Calculate the GDP per capita of Central African Republic.
10. Explain briefly whether each of the following would cause GDP to overstate or understate the degree of change in the broad standard of living.
 - a. The environment becomes dirtier
 - b. The crime rate declines
 - c. A greater variety of goods become available to consumers
 - d. Infant mortality declines

Review Questions

11. What are the main components of measuring GDP with what is demanded?
12. What are the main components of measuring GDP with what is produced?
13. Would you usually expect GDP as measured by what is demanded to be greater than GDP measured by what is supplied, or the reverse?
14. Why must you avoid double counting when measuring GDP?
15. What is the difference between a series of economic data over time measured in nominal terms versus the same data series over time measured in real terms?
16. How do you convert a series of nominal economic data over time to real terms?
17. What are typical GDP patterns for a high-income economy like the United States in the long run and the short run?
18. What are the two main difficulties that arise in comparing different countries's GDP?
19. List some of the reasons why economists should not consider GDP an effective measure of the standard of living in a country.

Critical Thinking Questions

20. U.S. macroeconomic data are among the best in the world. Given what you learned in the [Clear It Up](#) "How do statisticians measure GDP?", does this surprise you, or does this simply reflect the complexity of a modern economy?
21. What does GDP not tell us about the economy?
22. Should people typically pay more attention to their real income or their nominal income? If you choose the latter, why would that make sense in today's world? Would your answer be the same for the 1970s?
23. Why do you suppose that U.S. GDP is so much higher today than 50 or 100 years ago?
24. Why do you think that GDP does not grow at a steady rate, but rather speeds up and slows down?
25. Cross country comparisons of GDP per capita typically use purchasing power parity equivalent exchange rates, which are a measure of the long run equilibrium value of an exchange rate. In fact, we used PPP equivalent exchange rates in this module. Why could using market exchange rates, which sometimes change dramatically in a short period of time, be misleading?
26. Why might per capita GDP be only an imperfect measure of a country's standard of living?

27. How might you measure a “green” GDP?

Problems

28. Last year, a small nation with abundant forests cut down \$200 worth of trees. It then turned \$100 worth of trees into \$150 worth of lumber. It used \$100 worth of that lumber to produce \$250 worth of bookshelves. Assuming the country produces no other outputs, and there are no other inputs used in producing trees, lumber, and bookshelves, what is this nation's GDP? In other words, what is the value of the final goods the nation produced including trees, lumber and bookshelves?
29. The “prime” interest rate is the rate that banks charge their best customers. Based on the nominal interest rates and inflation rates in [Table 19.10](#), in which of the years would it have been best to be a lender? Based on the nominal interest rates and inflation rates in [Table 19.10](#), in which of the years given would it have been best to be a borrower?

Year	Prime Interest Rate	Inflation Rate
1970	7.9%	5.7%
1974	10.8%	11.0%
1978	9.1%	7.6%
1981	18.9%	10.3%

TABLE 19.10

30. A mortgage loan is a loan that a person makes to purchase a house. [Table 19.11](#) provides a list of the mortgage interest rate for several different years and the rate of inflation for each of those years. In which years would it have been better to be a person borrowing money from a bank to buy a home? In which years would it have been better to be a bank lending money?

Year	Mortgage Interest Rate	Inflation Rate
1984	12.4%	4.3%
1990	10%	5.4%
2001	7.0%	2.8%

TABLE 19.11

31. Ethiopia has a GDP of \$8 billion (measured in U.S. dollars) and a population of 55 million. Costa Rica has a GDP of \$9 billion (measured in U.S. dollars) and a population of 4 million. Calculate the per capita GDP for each country and identify which one is higher.
32. In 1980, Denmark had a GDP of \$70 billion (measured in U.S. dollars) and a population of 5.1 million. In 2000, Denmark had a GDP of \$160 billion (measured in U.S. dollars) and a population of 5.3 million. By what percentage did Denmark's GDP per capita rise between 1980 and 2000?
33. The Czech Republic has a GDP of 1,800 billion koruny. The exchange rate is 25 koruny/U.S. dollar. The Czech population is 20 million. What is the GDP per capita of the Czech Republic expressed in U.S. dollars?



FIGURE 20.1 Average Daily Calorie Consumption Not only has the number of calories that people consume per day increased, so has the amount of food calories that people are able to afford based on their working wages. (Credit: modification of "Daily Calorie Intake" by Lauren Manning/Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- The Relatively Recent Arrival of Economic Growth
- Labor Productivity and Economic Growth
- Components of Economic Growth
- Economic Convergence

Introduction to Economic Growth



BRING IT HOME

Calories and Economic Growth

On average, humans need about 2,500 calories a day to survive, depending on height, weight, and gender. The economist Brad DeLong estimates that the average worker in the early 1600s earned wages that could afford him 2,500 food calories. This worker lived in Western Europe. Two hundred years later, that same worker could afford 3,000 food calories. However, between 1800 and 1875, just a time span of just 75 years, economic growth was so rapid that western European workers could purchase 5,000 food calories a day. By 2012, a low skilled worker in an affluent Western European/North American country could afford to purchase 2.4 million food calories per day.

What caused such a rapid rise in living standards between 1800 and 1875 and thereafter? Why is it that many countries, especially those in Western Europe, North America, and parts of East Asia, can feed their populations more than adequately, while others cannot? We will look at these and other questions as we examine long-run economic growth.

Every country worries about economic growth. In the United States and other high-income countries, the question is whether economic growth continues to provide the same remarkable gains in our standard of living as it did during the twentieth century. Meanwhile, can middle-income countries like Brazil, Egypt, or Poland catch up to the higher-income countries, or must they remain in the second tier of per capita income? Of the world's population of roughly 7.5 billion people, about 1.1 billion are scraping by on incomes that average less than \$2 per day, not that different from the standard of living 2,000 years ago. Can the world's poor be lifted from their fearful poverty? As the 1995 Nobel laureate in economics, Robert E. Lucas Jr., once noted: "The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else."

Dramatic improvements in a nation's standard of living are possible. After the Korean War in the late 1950s, the Republic of Korea, often called South Korea, was one of the poorest economies in the world. Most South Koreans worked in peasant agriculture. According to the British economist Angus Maddison, who devoted life's work to measuring GDP and population in the world economy, GDP per capita in 1990 international dollars was \$854 per year. From the 1960s to the early twenty-first century, a time period well within the lifetime and memory of many adults, the South Korean economy grew rapidly. Over these four decades, GDP per capita increased by more than 6% per year. According to the World Bank, GDP for South Korea now exceeds \$30,000 in nominal terms, placing it firmly among high-income countries like Italy, New Zealand, and Israel. Measured by total GDP in 2015, South Korea is the eleventh-largest economy in the world. For a nation of 50 million people, this transformation is extraordinary.

South Korea is a standout example, but it is not the only case of rapid and sustained economic growth. Other East Asian nations, like Thailand and Indonesia, have seen very rapid growth as well. China has grown enormously since it enacted market-oriented economic reforms around 1980. GDP per capita in high-income economies like the United States also has grown dramatically albeit over a longer time frame. Since the Civil War, the U.S. economy has transformed from a primarily rural and agricultural economy to an economy based on services, manufacturing, and technology.

20.1 The Relatively Recent Arrival of Economic Growth

LEARNING OBJECTIVES

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

- Explain the conditions that have allowed for modern economic growth in the last two centuries
- Analyze the influence of public policies on an economy's long-run economic growth

Let's begin with a brief overview of spectacular economic growth patterns around the world in the last two centuries. We commonly refer to this as the period of **modern economic growth**. (Later in the chapter we will discuss lower economic growth rates and some key ingredients for economic progress.) Rapid and sustained economic growth is a relatively recent experience for the human race. Before the last two centuries, although rulers, nobles, and conquerors could afford some extravagances and although economies rose above the subsistence level, the average person's standard of living had not changed much for centuries.

Progressive, powerful economic and institutional changes started to have a significant effect in the late eighteenth and early nineteenth centuries. According to the Dutch economic historian Jan Luiten van Zanden, slavery-based societies, favorable demographics, global trading routes, and standardized trading institutions that spread with different empires set the stage for the Industrial Revolution to succeed. The **Industrial Revolution** refers to the widespread use of power-driven machinery and the economic and social changes that resulted in the first half of the 1800s. Ingenious machines—the steam engine, the power loom, and the steam locomotive—performed tasks that otherwise would have taken vast numbers of workers to do. The Industrial Revolution began in Great Britain, and soon spread to the United States, Germany, and other countries.

The jobs for ordinary people working with these machines were often dirty and dangerous by modern standards, but the alternative jobs of that time in peasant agriculture and small-village industry were often

dirty and dangerous, too. The new jobs of the Industrial Revolution typically offered higher pay and a chance for social mobility. A self-reinforcing cycle began: New inventions and investments generated profits, the profits provided funds for more new investment and inventions, and the investments and inventions provided opportunities for further profits. Slowly, a group of national economies in Europe and North America emerged from centuries of sluggishness into a period of rapid modern growth. During the last two centuries, the average GDP growth rate per capita in the leading industrialized countries has been about 2% per year. What were times like before then? Read the following Clear It Up feature for the answer.



CLEAR IT UP

What were economic conditions like before 1870?

Angus Maddison, a quantitative economic historian, led the most systematic inquiry into national incomes before 1870. Economists recently have refined and used his methods to compile GDP per capita estimates from year 1 C.E. to 1348. [Table 20.1](#) is an important counterpoint to most of the narrative in this chapter. It shows that nations can decline as well as rise. A wide array of forces, such as epidemics, natural and weather-related disasters, the inability to govern large empires, and the remarkably slow pace of technological and institutional progress explain declines in income. Institutions are the traditions and laws by which people in a community agree to behave and govern themselves. Such institutions include marriage, religion, education, and laws of governance. Institutional progress is the development and codification of these institutions to reinforce social order, and thus, economic growth.

One example of such an institution is the Magna Carta (Great Charter), which the English nobles forced King John to sign in 1215. The Magna Carta codified the principles of due process, whereby a free man could not be penalized unless his peers had made a lawful judgment against him. The United States in its own constitution later adopted this concept. This social order may have contributed to England's GDP per capita in 1348, which was second to that of northern Italy.

In studying economic growth, a country's institutional framework plays a critical role. [Table 20.1](#) also shows relative global equality for almost 1,300 years. After this, we begin to see significant divergence in income (not in the table).

Year	Northern Italy	Spain	England	Holland	Byzantium	Iraq	Egypt	Japan
1	\$800	\$600	\$600	\$600	\$700	\$700	\$700	-
730	-	-	-	-	-	\$920	\$730	\$402
1000	-	-	-	-	\$600	\$820	\$600	-
1150	-	-	-	-	\$580	\$680	\$660	\$520
1280	-	-	-	-	-	-	\$670	\$527
1300	\$1,588	\$864	\$892	-	-	-	\$610	-
1348	\$1,486	\$907	\$919	-	-	-	-	-

TABLE 20.1 GDP Per Capita Estimates in Current International Dollars from AD 1 to 1348
(Source: Bolt and van Zanden. "The First Update of the Maddison Project. Re-Estimating Growth Before 1820." 2013)

Another fascinating and underreported fact is the high levels of income, compared to others at that time, attained by the Islamic Empire Abbasid Caliphate—which was founded in present-day Iraq in 730 C.E. At its height, the empire spanned large regions of the Middle East, North Africa, and Spain until its gradual decline over 200 years.

The Industrial Revolution led to increasing inequality among nations. Some economies took off, whereas others, like many of those in Africa or Asia, remained close to a subsistence standard of living. General calculations show that the 17 countries of the world with the most-developed economies had, on average, 2.4 times the GDP per capita of the world's poorest economies in 1870. By 1960, the most developed economies had 4.2 times the GDP per capita of the poorest economies.

However, by the middle of the twentieth century, some countries had shown that catching up was possible. Japan's economic growth took off in the 1960s and 1970s, with a growth rate of real GDP per capita averaging 11% per year during those decades. Certain countries in Latin America experienced a boom in economic growth in the 1960s as well. In Brazil, for example, GDP per capita expanded by an average annual rate of 11.1% from 1968 to 1973. In the 1970s, some East Asian economies, including South Korea, Thailand, and Taiwan, saw rapid growth. In these countries, growth rates of 11% to 12% per year in GDP per capita were not uncommon. More recently, China, with its population of nearly 1.4 billion people, grew at a per capita rate 9% per year from 1984 into the 2000s and still average high rates of growth (more than 5% today). India, with a population of 1.4 billion, has shown promising signs of economic growth, with growth in GDP per capita of about 4% per year during the 1990s and climbing toward 7% to 8% per year in the 2000s and 2010s.

LINK IT UP

Visit this [website \(http://openstax.org/l/asiadevbank\)](http://openstax.org/l/asiadevbank) to read about the Asian Development Bank.

These waves of catch-up economic growth have not reached all shores. In certain African countries like Niger, Tanzania, and Sudan, for example, GDP per capita at the start of the 2000s was still less than \$300, not much higher than it was in the nineteenth century and for centuries before that. In the context of the overall situation of low-income people around the world, the good economic news from China (population: 1.4 billion) and India (population: 1.3 billion) is, nonetheless, astounding and heartening.

Economic growth in the last two centuries has made a striking change in the human condition. Richard Easterlin, an economist at the University of Southern California, wrote in 2000:

By many measures, a revolution in the human condition is sweeping the world. Most people today are better fed, clothed, and housed than their predecessors two centuries ago. They are healthier, live longer, and are better educated. Women's lives are less centered on reproduction and political democracy has gained a foothold. Although Western Europe and its offshoots have been the leaders of this advance, most of the less developed nations have joined in during the 20th century, with the newly emerging nations of sub-Saharan Africa the latest to participate. Although the picture is not one of universal progress, it is the greatest advance in the human condition of the world's population ever achieved in such a brief span of time.

Rule of Law and Economic Growth

Economic growth depends on many factors. Key among those factors is adherence to the **rule of law** and protection of property rights and **contractual rights** by a country's government so that markets can work effectively and efficiently. Laws must be clear, public, fair, enforced, and equally applicable to all members of society. Property rights, as you might recall from [Environmental Protection and Negative Externalities](#) are the rights of individuals and firms to own property and use it as they see fit. If you have \$100, you have the right to use that money, whether you spend it, lend it, or keep it in a jar. It is your property. The definition of property includes physical property as well as the right to your training and experience, especially since your training is what determines your livelihood. Using this property includes the right to enter into contracts with other parties with your property. Individuals or firms must own the property to enter into a contract.

Contractual rights, then, are based on property rights and they allow individuals to enter into agreements with others regarding the use of their property providing recourse through the legal system in the event of

noncompliance. One example is the employment agreement: a skilled surgeon operates on an ill person and expects payment. Failure to pay would constitute property theft by the patient. The theft is property the services that the surgeon provided. In a society with strong property rights and contractual rights, the terms of the patient–surgeon contract will be fulfilled, because the surgeon would have recourse through the court system to extract payment from that individual. Without a legal system that enforces contracts, people would not be likely to enter into contracts for current or future services because of the risk of non-payment. This would make it difficult to transact business and would slow economic growth.

The World Bank considers a country’s legal system effective if it upholds property rights and contractual rights. The World Bank has developed a ranking system for countries’ legal systems based on effective protection of property rights and rule-based governance using a scale from 1 to 6, with 1 being the lowest and 6 the highest rating. In 2020, the world average ranking was 2.9. The three countries with the lowest ranking of 1.0 were Somalia and Eritrea, with South Sudan at 1.5. Their GDP per capita was \$875, \$1,625, and \$1,234.70 respectively. The World Bank also cites Afghanistan (GDP per capita \$2,087.60) as having a low standard of living, weak government structure, and lack of adherence to the rule of law, which has stymied its economic growth. The landlocked Central African Republic (GDP per capita \$979.60) has poor economic resources as well as political instability and is a source of children used in human trafficking. Zimbabwe (GDP per capita \$2,895.40) has had declining and often negative growth for much of the period since 1998. Land redistribution and price controls have disrupted the economy, and corruption and violence have dominated the political process. Although global economic growth has increased, those countries lacking a clear system of property rights and an independent court system free from corruption have lagged far behind.

20.2 Labor Productivity and Economic Growth

LEARNING OBJECTIVES

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

- Identify the role of labor productivity in promoting economic growth
- Analyze the sources of economic growth using the aggregate production function
- Measure an economy’s rate of productivity growth
- Evaluate the power of sustained growth

Sustained long-term economic growth comes from increases in worker productivity, which essentially means how well we do things. In other words, how efficient is your nation with its time and workers? **Labor productivity** is the value that each employed person creates per unit of their input. The easiest way to comprehend labor productivity is to imagine a Canadian worker who can make 10 loaves of bread in an hour versus a U.S. worker who in the same hour can make only two loaves of bread. In this fictional example, the Canadians are more productive. More productivity essentially means you can do more in the same amount of time. This in turn frees up resources for workers to use elsewhere.

What determines how productive workers are? The answer is pretty intuitive. The first determinant of labor productivity is human capital. **Human capital** is the accumulated knowledge (from education and experience), skills, and expertise that the average worker in an economy possesses. Typically the higher the average level of education in an economy, the higher the accumulated human capital and the higher the labor productivity.

The second factor that determines labor productivity is technological change. **Technological change** is a combination of **invention**—advances in knowledge—and **innovation**, which is putting those advances to use in a new product or service. For example, the transistor was invented in 1947. It allowed us to miniaturize the footprint of electronic devices and use less power than the tube technology that came before it. Innovations since then have produced smaller and better transistors that are ubiquitous in products as varied as smart-phones, computers, and escalators. Developing the transistor has allowed workers to be anywhere with smaller devices. People can use these devices to communicate with other workers, measure product quality or do any other task in less time, improving worker productivity.

The third factor that determines labor productivity is economies of scale. Recall that economies of scale are the cost advantages that industries obtain due to size. (Read more about economies of scale in [Production, Cost and Industry Structure](#).) Consider again the case of the fictional Canadian worker who could produce 10 loaves of bread in an hour. If this difference in productivity was due only to economies of scale, it could be that the Canadian worker had access to a large industrial-size oven while the U.S. worker was using a standard residential size oven.

Now that we have explored the determinants of worker productivity, let's turn to how economists measure economic growth and productivity.

Sources of Economic Growth: The Aggregate Production Function

To analyze the sources of economic growth, it is useful to think about a **production function**, which is the technical relationship by which economic inputs like labor, machinery, and raw materials are turned into outputs like goods and services that consumers use. A microeconomic production function describes a firm's or perhaps an industry's inputs and outputs. In macroeconomics, we call the connection from inputs to outputs for the entire economy an **aggregate production function**.

Components of the Aggregate Production Function

Economists construct different production functions depending on the focus of their studies. [Figure 20.2](#) presents two examples of aggregate production functions. In the first production function in [Figure 20.2](#) (a), the output is GDP. The inputs in this example are workforce, human capital, physical capital, and technology. We discuss these inputs further in the module, Components of Economic Growth.

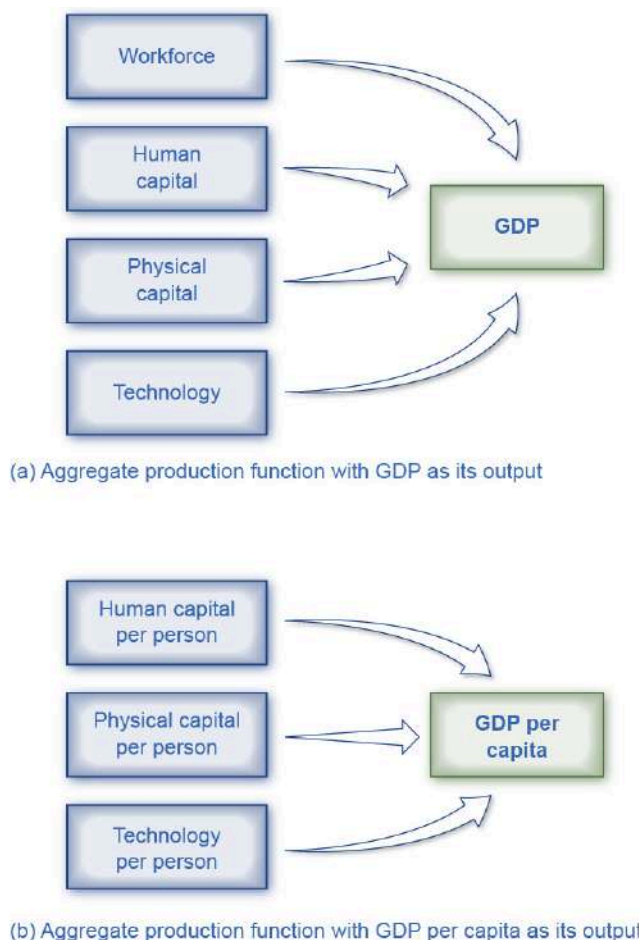


FIGURE 20.2 Aggregate Production Functions An aggregate production function shows what goes into producing

the output for an overall economy. (a) This aggregate production function has GDP as its output. (b) This aggregate production function has GDP per capita as its output. Because we calculate it on a per-person basis, we already figure the labor input into the other factors and we do not need to list it separately.

Measuring Productivity

An economy's rate of productivity growth is closely linked to the growth rate of its GDP per capita, although the two are not identical. For example, if the percentage of the population who holds jobs in an economy increases, GDP per capita will increase but the productivity of individual workers may not be affected. Over the long term, the only way that GDP per capita can grow continually is if the productivity of the average worker rises or if there are complementary increases in capital.

A common measure of U.S. productivity per worker is dollar value per hour the worker contributes to the employer's output. This measure excludes government workers, because their output is not sold in the market and so their productivity is hard to measure. It also excludes farming, which accounts for only a relatively small share of the U.S. economy. Figure 20.3 shows an index of output per hour, with 2012 as the base year (when the index equals 100). The index equaled 110.5 in 2020. In 1977, the index equaled about 50, which shows that workers have more than doubled their productivity since then.

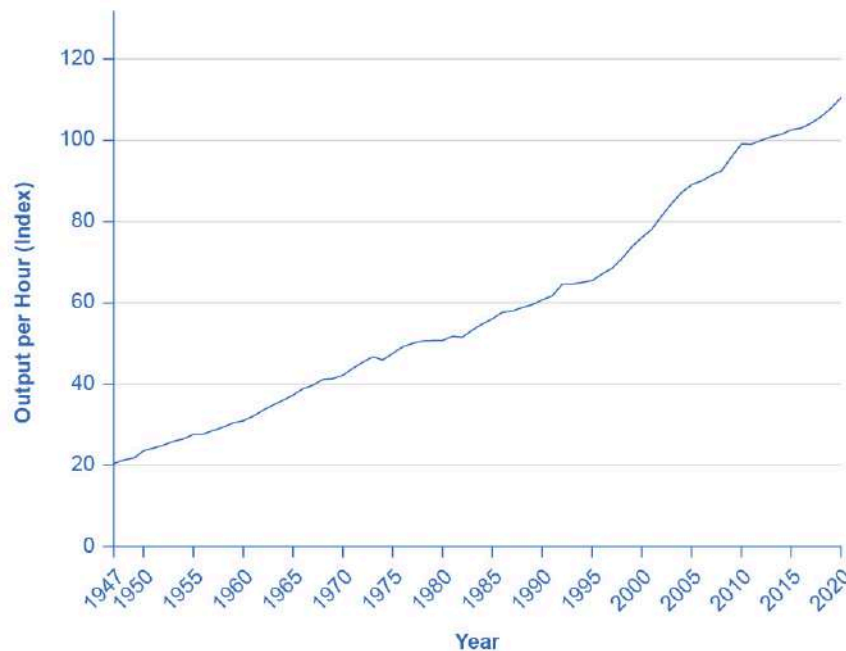


FIGURE 20.3 Output per Hour Worked in the U.S. Economy, 1947–2020 Output per hour worked is a measure of worker productivity. In the U.S. economy, worker productivity rose more quickly in the 1960s and the mid-1990s compared with the 1970s and 1980s. However, these growth-rate differences are only a few percentage points per year. Look carefully to see them in the changing slope of the line. The average U.S. worker produced over twice as much per hour in 2020 than they did in the 1970s. (Source: U.S. Department of Labor, Bureau of Labor Statistics.)

[Click to view content \(https://openstax.org/books/principles-economics-3e/pages/20-2-labor-productivity-and-economic-growth\)](https://openstax.org/books/principles-economics-3e/pages/20-2-labor-productivity-and-economic-growth)

A graph has an X-axis with years progressing from 1955 to 2020 and a Y axis labeled Percent Change at Annual Rate. The graphed data moves up and down across a zero line indicating change year over year. In 1970, 1974, 1981, 1983, 2008, and 2020, the rate was quite low, as the U.S. was undergoing recessions.

According to the Department of Labor, U.S. productivity growth was fairly strong in the 1950s but then declined in the 1970s and 1980s before rising again in the second half of the 1990s and the first half of the 2000s. In fact, the rate of productivity measured by the change in output per hour worked averaged 2.8% per

year from 1947 to 1973; dropped to 1.2% per year from 1973 to 1979; increased to 1.5% per year from 1979 to 1990; increased again to 2.2% from 1990 to 2000; increased even more to 2.7% from 2000 to 2007; and then decreased to 1.4% from 2007 to 2020. [Figure 20.4](#) shows average annual rates of productivity growth averaged over time since 1947.

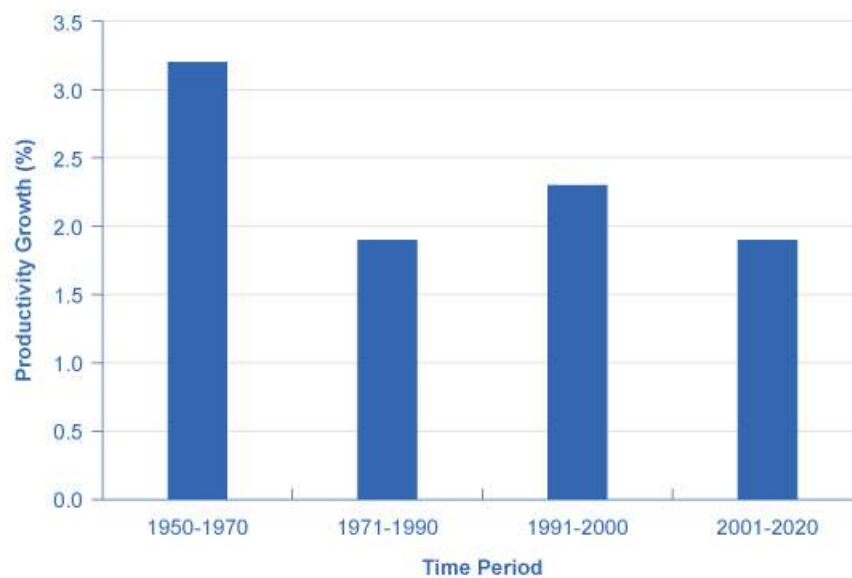


FIGURE 20.4 Productivity Growth Since 1947 U.S. growth in worker productivity was very high between 1947 and 1973. It then declined to lower levels in the later 1970s and the 1980s. The late 1990s and early 2000s saw productivity rebound, but then productivity sagged a bit between 2001 and 2020. Some think the productivity rebound of the late 1990s and early 2000s marks the start of a “new economy” built on higher productivity growth, but we cannot determine this until more time has passed. (Source: U.S. Department of Labor, Bureau of Labor Statistics.)

The “New Economy” Controversy

In recent years a controversy has been brewing among economists about the resurgence of U.S. productivity in the second half of the 1990s. One school of thought argues that the United States had developed a “new economy” based on the extraordinary advances in communications and information technology of the 1990s. The most optimistic proponents argue that it would generate higher average productivity growth for decades to come. The pessimists, alternatively, argue that even five or ten years of stronger productivity growth does not prove that higher productivity will last for the long term. It is hard to infer anything about long-term productivity trends during the later part of the 2000s, because the steep 2008-2009 recession, with its sharp but not completely synchronized declines in output and employment, complicates any interpretation. While productivity growth was high in 2009 and 2010 (around 3%), it has slowed down over the last decade.

Productivity growth is also closely linked to the average level of wages. Over time, the amount that firms are willing to pay workers will depend on the value of the output those workers produce. If a few employers tried to pay their workers less than what those workers produced, then those workers would receive offers of higher wages from other profit-seeking employers. If a few employers mistakenly paid their workers more than what those workers produced, those employers would soon end up with losses. In the long run, productivity per hour is the most important determinant of the average wage level in any economy. To learn how to compare economies in this regard, follow the steps in the following Work It Out feature.

WORK IT OUT

Comparing the Economies of Two Countries

The Organization for Economic Co-operation and Development (OECD) tracks data on the annual growth rate of real GDP per hour worked. You can find these data on the OECD data webpage “Growth in GDP per capita, productivity and ULC” at [this \(http://stats.oecd.org/Index.aspx?DataSetCode=PDB_GR\)](http://stats.oecd.org/Index.aspx?DataSetCode=PDB_GR) website.

Step 1. Visit the OECD website given above and select two countries to compare.

Step 2. On the drop-down menu “Subject,” select “GDP per capita, constant prices,” and under “Measure,” select “Annual growth/change.” Then record the data for the countries you have chosen for the five most recent years.

Step 3. Go back to the drop-down “Subject” menu and select “GDP per hour worked, constant prices,” and under “Measure” again select “Annual growth/change.” Select data for the same years for which you selected GDP per capita data.

Step 4. Compare real GDP growth for both countries. [Table 20.2](#) provides an example of a comparison between Australia and Belgium.

Australia	2011	2012	2013	2014	2015
Real GDP/Capita Growth (%)	2.3%	1.5%	1.3%	1.4	0.1%
Real GDP Growth/Hours Worked (%)	1.7%	−0.1%	1.4%	2.2%	−0.2%
Belgium	2011	2012	2013	2014	2015
Real GDP/Capita Growth (%)	0.9	−0.6	−0.5	1.2	1.0
Real GDP Growth/Hours Worked (%)	−0.5	−0.3	0.4	1.4	0.9

TABLE 20.2

Step 5. For both measures, growth in Australia is greater than growth in Belgium for the first four years. In addition, there are year-to-year fluctuations. Many factors can affect growth. For example, one factor that may have contributed to Australia's stronger growth may be its larger inflows of immigrants, who generally contribute to economic growth.

The Power of Sustained Economic Growth

Nothing is more important for people's standard of living than sustained economic growth. Even small changes in the rate of growth, when sustained and compounded over long periods of time, make an enormous difference in the standard of living. Consider [Table 20.3](#), in which the rows of the table show several different rates of growth in GDP per capita and the columns show different periods of time. Assume for simplicity that an economy starts with a GDP per capita of 100. The table then applies the following formula to calculate what GDP will be at the given growth rate in the future:

$$\text{GDP at starting date} \times (1 + \text{growth rate of GDP})^{\text{years}} = \text{GDP at end date}$$

For example, an economy that starts with a GDP of 100 and grows at 3% per year will reach a GDP of 209 after 25 years; that is, $100 (1.03)^{25} = 209$.

The slowest rate of GDP per capita growth in the table, just 1% per year, is similar to what the United States

experienced during its weakest years of productivity growth. The second highest rate, 3% per year, is close to what the U.S. economy experienced during the strong economy of the late 1990s and into the 2000s. Higher rates of per capita growth, such as 5% or 8% per year, represent the experience of rapid growth in economies like Japan, Korea, and China.

[Table 20.3](#) shows that even a few percentage points of difference in economic growth rates will have a profound effect if sustained and compounded over time. For example, an economy growing at a 1% annual rate over 50 years will see its GDP per capita rise by a total of 64%, from 100 to 164 in this example. However, a country growing at a 5% annual rate will see (almost) the same amount of growth—from 100 to 163—over just 10 years. Rapid rates of economic growth can bring profound transformation. (See the following Clear It Up feature on the relationship between compound growth rates and compound interest rates.) If the rate of growth is 8%, young adults starting at age 20 will see the average standard of living in their country more than double by the time they reach age 30, and grow more than sixfold by the time they reach age 45.

Growth Rate	Value of an original 100 in 10 Years	Value of an original 100 in 25 Years	Value of an original 100 in 50 Years
1%	110	128	164
3%	134	209	438
5%	163	339	1,147
8%	216	685	4,690

TABLE 20.3 Growth of GDP over Different Time Horizons



CLEAR IT UP

How are compound growth rates and compound interest rates related?

The formula for GDP growth rates over different periods of time, as [Figure 20.3](#) shows, is exactly the same as the formula for how a given amount of financial savings grows at a certain interest rate over time, as presented in [Choice in a World of Scarcity](#). Both formulas have the same ingredients:

- an original starting amount, in one case GDP and in the other case an amount of financial saving;
- a percentage increase over time, in one case the GDP growth rate and in the other case an interest rate;
- and an amount of time over which this effect happens.

Recall that compound interest is interest that is earned on past interest. It causes the total amount of financial savings to grow dramatically over time. Similarly, compound rates of economic growth, or the **compound growth rate**, means that we multiply the rate of growth by a base that includes past GDP growth, with dramatic effects over time.

For example, in 2020, the Central Intelligence Agency's World Fact Book reported that South Korea had a GDP of \$2.2 trillion. With a growth rate of 2.8% per year, South Korea's GDP will be \$2.5 trillion in five years. If we apply the growth rate to each year's ending GDP for the next five years, we will calculate that at the end of year one, GDP is \$2.3 trillion. In year two, we start with the end-of-year one value of \$1.72 and increase it by 2.8%. Year three starts with the end-of-year two GDP, and we increase it by 2.8% and so on, as [Table 20.4](#) depicts.

Year	Starting GDP	Growth Rate 2.8%	Year-End Amount
1	\$2.2 Trillion ×	(1+0.028)	\$2.3 Trillion
2	\$2.3 Trillion ×	(1+0.028)	\$2.3 Trillion
3	\$2.3 Trillion ×	(1+0.028)	\$2.4 Trillion
4	\$2.4 Trillion ×	(1+0.028)	\$2.5 Trillion
5	\$2.5 Trillion ×	(1+0.028)	\$2.5 Trillion

TABLE 20.4

Another way to calculate the growth rate is to apply the following formula:

$$\text{Future Value} = \text{Present Value} \times (1 + g)^n$$

Where “future value” is the value of GDP five years hence, “present value” is the starting GDP amount of \$1.67 trillion, “g” is the growth rate of 2.8%, and “n” is the number of periods for which we are calculating growth.

$$\text{Future Value} = 1.67 \times (1+0.028)^5 = \$1.92 \text{ trillion}$$

20.3 Components of Economic Growth

LEARNING OBJECTIVES

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

- Discuss the components of economic growth, including physical capital, human capital, and technology
- Explain capital deepening and its significance
- Analyze the methods employed in economic growth accounting studies
- Identify factors that contribute to a healthy climate for economic growth

Over decades and generations, seemingly small differences of a few percentage points in the annual rate of economic growth make an enormous difference in GDP per capita. In this module, we discuss some of the components of economic growth, including physical capital, human capital, and technology.

The category of **physical capital** includes the plant and equipment that firms use as well as things like roads (also called **infrastructure**). Again, greater physical capital implies more output. Physical capital can affect productivity in two ways: (1) an increase in the *quantity* of physical capital (for example, more computers of the same quality); and (2) an increase in the *quality* of physical capital (same number of computers but the computers are faster, and so on). **Human capital** refers to the skills and knowledge that make workers productive. Human capital and physical capital accumulation are similar: In both cases, investment now pays off in higher productivity in the future.

The category of **technology** is the “joker in the deck.” Earlier we described it as the combination of invention and innovation. When most people think of new technology, the invention of new products like the laser, the smartphone, or some new wonder drug come to mind. In food production, developing more drought-resistant seeds is another example of technology. Technology, as economists use the term, however, includes still more. It includes new ways of organizing work, like the invention of the assembly line, new methods for ensuring better quality of output in factories, and innovative institutions that facilitate the process of converting inputs into output. In short, technology comprises all the advances that make the existing machines and other inputs produce more, and at higher quality, as well as altogether new products.

It may not make sense to compare the GDPs of China and say, Benin, simply because of the great difference in population size. To understand economic growth, which is really concerned with the growth in living

standards of an average person, it is often useful to focus on GDP per capita. Using GDP per capita also makes it easier to compare countries with smaller numbers of people, like Belgium, Uruguay, or Zimbabwe, with countries that have larger populations, like the United States, the Russian Federation, or Nigeria.

To obtain a per capita production function, divide each input in [Figure 20.2\(a\)](#) by the population. This creates a second aggregate production function where the output is GDP per capita (that is, GDP divided by population). The inputs are the average level of human capital per person, the average level of physical capital per person, and the level of technology per person—see [Figure 20.2\(b\)](#). The result of having population in the denominator is mathematically appealing. Increases in population lower per capita income. However, increasing population is important for the average person only if the rate of income growth exceeds population growth. A more important reason for constructing a per capita production function is to understand the contribution of human and physical capital.

Capital Deepening

When society increases the level of capital per person, we call the result **capital deepening**. The idea of capital deepening can apply both to additional human capital per worker and to additional physical capital per worker.

Recall that one way to measure human capital is to look at the average levels of education in an economy. [Figure 20.5](#) illustrates the human capital deepening for U.S. workers by showing that the proportion of the U.S. population with a high school and a college degree is rising. As recently as 1970, for example, only about half of U.S. adults had at least a high school diploma. By the start of the twenty-first century, more than 80% of adults had graduated from high school. The idea of human capital deepening also applies to the years of experience that workers have, but the average experience level of U.S. workers has not changed much in recent decades. Thus, the key dimension for deepening human capital in the U.S. economy focuses more on additional education and training than on a higher average level of work experience.

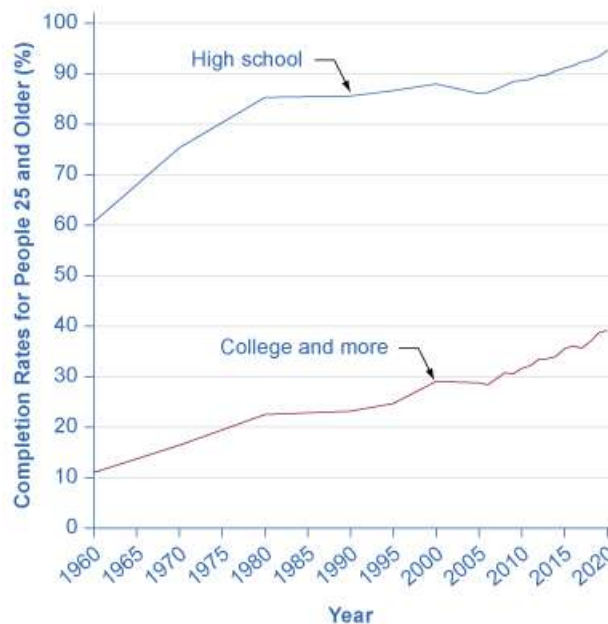


FIGURE 20.5 Human Capital Deepening in the U.S. Rising levels of education for persons 25 and older show the deepening of human capital in the U.S. economy. Even today, under one-third of U.S. adults have completed a four-year college degree. There is clearly room for additional deepening of human capital to occur. (Source: Penn World Tables, 10.0 <https://www.rug.nl/ggdc/productivity/pwt/?lang=en>)

[Figure 20.6](#) shows physical capital deepening in the U.S. economy. The average U.S. worker in the late 2000s was working with physical capital worth almost three times as much as that of the average worker of the early

1950s.



FIGURE 20.6 Physical Capital per Worker in the United States The value of the physical capital, measured by plant and equipment, used by the average worker in the U.S. economy has risen over the decades. The increase may have leveled off a bit in the 1970s and 1980s, which were, not coincidentally, times of slower-than-usual growth in worker productivity. We see a renewed increase in physical capital per worker in the late 1990s, followed by a flattening in the early 2000s. (Source: Center for International Comparisons of Production, Income and Prices, University of Pennsylvania)

Not only does the current U.S. economy have better-educated workers with more and improved physical capital than it did several decades ago, but these workers have access to more advanced technologies. Growth in technology is impossible to measure with a simple line on a graph, but evidence that we live in an age of technological marvels is all around us—discoveries in genetics and in the structure of particles, the wireless internet, and other inventions almost too numerous to count. The U.S. Patent and Trademark Office typically has issued more than 150,000 patents annually in recent years.

This recipe for economic growth—investing in labor productivity, with investments in human capital and technology, as well as increasing physical capital—also applies to other economies. South Korea, for example, already achieved universal enrollment in primary school (the equivalent of kindergarten through sixth grade in the United States) by 1965, when Korea’s GDP per capita was still near its rock bottom low. By the late 1980s, Korea had achieved almost universal secondary school education (the equivalent of a high school education in the United States). With regard to physical capital, Korea’s rates of investment had been about 15% of GDP at the start of the 1960s, but doubled to 30–35% of GDP by the late 1960s and early 1970s. With regard to technology, South Korean students went to universities and colleges around the world to obtain scientific and technical training, and South Korean firms reached out to study and form partnerships with firms that could offer them technological insights. These factors combined to foster South Korea’s high rate of economic growth.

Growth Accounting Studies

Since the late 1950s, economists have conducted growth accounting studies to determine the extent to which

physical and human capital deepening and technology have contributed to growth. The usual approach uses an aggregate production function to estimate how much of per capita economic growth can be attributed to growth in physical capital and human capital. We can measure these two inputs at least roughly. The part of growth that is unexplained by measured inputs, called the residual, is then attributed to growth in technology. The exact numerical estimates differ from study to study and from country to country, depending on how researchers measured these three main factors and over what time horizons. For studies of the U.S. economy, three lessons commonly emerge from growth accounting studies.

First, technology is typically the most important contributor to U.S. economic growth. Growth in human capital and physical capital often explains only half or less than half of the economic growth that occurs. New ways of doing things are tremendously important.

Second, while investment in physical capital is essential to growth in labor productivity and GDP per capita, building human capital is at least as important. Economic growth is not just a matter of more machines and buildings. One vivid example of the power of human capital and technological knowledge occurred in Europe in the years after World War II (1939–1945). During the war, a large share of Europe's physical capital, such as factories, roads, and vehicles, was destroyed. Europe also lost an overwhelming amount of human capital in the form of millions of men, women, and children who died during the war. However, the powerful combination of skilled workers and technological knowledge, working within a market-oriented economic framework, rebuilt Europe's productive capacity to an even higher level within less than two decades.

A third lesson is that these three factors of human capital, physical capital, and technology work together. Workers with a higher level of education and skills are often better at coming up with new technological innovations. These technological innovations are often ideas that cannot increase production until they become a part of new investment in physical capital. New machines that embody technological innovations often require additional training, which builds worker skills further. If the recipe for economic growth is to succeed, an economy needs all the ingredients of the aggregate production function. See the following Clear It Up feature for an example of how human capital, physical capital, and technology can combine to significantly impact lives.



CLEAR IT UP

How do girls' education and economic growth relate in low-income countries?

In the early 2000s, according to the World Bank, about 110 million children between the ages of 6 and 11 were not in school—and about two-thirds of them were girls. In Afghanistan, for example, the literacy rate for those aged 15–24 for the period 2005–2014 was 62% for males and only 32% for females. In Benin, in West Africa, it was 55% for males and 31% for females. In Nigeria, Africa's most populous country, it was 76% for males and 58 percent for females.

Whenever any child does not receive a basic education, it is both a human and an economic loss. In low-income countries, wages typically increase by an average of 10 to 20% with each additional year of education. There is, however, some intriguing evidence that helping girls in low-income countries to close the education gap with boys may be especially important, because of the social role that many of the girls will play as mothers and homemakers.

Girls in low-income countries who receive more education tend to grow up to have fewer, healthier, better-educated children. Their children are more likely to be better nourished and to receive basic health care like immunizations. Economic research on women in low-income economies backs up these findings. When 20 women obtain one additional year of schooling, as a group they will, on average, have one less child. When 1,000 women obtain one additional year of schooling, on average one to two fewer women from that group will die in childbirth. When a woman stays in school an additional year, that factor alone means that, on average, each of her children will spend an additional half-year in school. Education for girls is a good investment because it is an investment in economic growth with benefits beyond the current generation.

A Healthy Climate for Economic Growth

While physical and human capital deepening and better technology are important, equally important to a nation's well-being is the climate or system within which these inputs are cultivated. Both the type of market economy and a legal system that governs and sustains property rights and contractual rights are important contributors to a healthy economic climate.

A healthy economic climate usually involves some sort of market orientation at the microeconomic, individual, or firm decision-making level. Markets that allow personal and business rewards and incentives for increasing human and physical capital encourage overall macroeconomic growth. For example, when workers participate in a competitive and well-functioning labor market, they have an incentive to acquire additional human capital, because additional education and skills will pay off in higher wages. Firms have an incentive to invest in physical capital and in training workers, because they expect to earn higher profits for their shareholders. Both individuals and firms look for new technologies, because even small inventions can make work easier or lead to product improvement. Collectively, such individual and business decisions made within a market structure add up to macroeconomic growth. Much of the rapid growth since the late nineteenth century has come from harnessing the power of competitive markets to allocate resources. This market orientation typically reaches beyond national borders and includes openness to international trade.

A general orientation toward markets does not rule out important roles for government. There are times when markets fail to allocate capital or technology in a manner that provides the greatest benefit for society as a whole. The government's role is to correct these failures. In addition, government can guide or influence markets toward certain outcomes. The following examples highlight some important areas that governments around the world have chosen to invest in to facilitate capital deepening and technology:

- **Education.** The Danish government requires all children under 16 to attend school. They can choose to attend a public school (*Folkeskole*) or a private school. Students do not pay tuition to attend *Folkeskole*. Thirteen percent of primary/secondary (elementary/high) school is private, and the government supplies vouchers to citizens who choose private school.
- **Savings and Investment.** In the United States, as in other countries, the government taxes gains from private investment. Low capital gains taxes encourage investment and so also economic growth.
- **Infrastructure.** The Japanese government in the mid-1990s undertook significant infrastructure projects to improve roads and public works. This in turn increased the stock of physical capital and ultimately economic growth.
- **Special Economic Zones.** The island of Mauritius is one of the few African nations to encourage international trade in government-supported **special economic zones (SEZ)**. These are areas of the country, usually with access to a port where, among other benefits, the government does not tax trade. As a result of its SEZ, Mauritius has enjoyed above-average economic growth since the 1980s. Free trade does not have to occur in an SEZ however. Governments can encourage international trade across the board, or surrender to protectionism.
- **Scientific Research.** The European Union has strong programs to invest in scientific research. The researchers Abraham García and Pierre Mohnen demonstrate that firms which received support from the Austrian government actually increased their research intensity and had more sales. Governments can support scientific research and technical training that helps to create and spread new technologies. Governments can also provide a legal environment that protects the ability of inventors to profit from their inventions.

There are many more ways in which the government can play an active role in promoting economic growth. We explore them in other chapters and in particular in [Macroeconomic Policy Around the World](#). A healthy climate for growth in GDP per capita and labor productivity includes human capital deepening, physical capital deepening, and technological gains, operating in a market-oriented economy with supportive government policies.

20.4 Economic Convergence

LEARNING OBJECTIVES

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

- Explain economic convergence
- Analyze various arguments for and against economic convergence
- Evaluate the speed of economic convergence between high-income countries and the rest of the world

Some low-income and middle-income economies around the world have shown a pattern of **convergence**, in which their economies grow faster than those of high-income countries. GDP increased by an average rate of 2.7% per year in the 1990s and 1.7% per year from 2010 to 2019 in the high-income countries of the world, which include the United States, Canada, the European Union countries, Japan, Australia, and New Zealand.

[Table 20.5](#) lists eight countries that belong to an informal “fast growth club.” These countries averaged GDP growth (after adjusting for inflation) of at least 5% per year in both the time periods from 1990 to 2000 and from 2010 to 2019. Since economic growth in these countries has exceeded the average of the world’s high-income economies, these countries may converge with the high-income countries. The second part of [Table 20.5](#) lists the “slow growth club,” which consists of countries that averaged GDP growth of 2% per year or less (after adjusting for inflation) during the same time periods. The final portion of [Table 20.5](#) shows GDP growth rates for the countries of the world divided by income.

Country	Average Growth Rate of Real GDP 1990–2000	Growth Rate of Real GDP 2010–2019
<i>Fast Growth Club (5% or more per year in both time periods)</i>		
Cambodia	7.1%	7.0%
China	10.6%	7.3%
India	6.0%	6.7%
Ireland	7.5%	6.3%
Laos	6.5%	7.3%
Mozambique	6.4%	5.6%
Uganda	7.1%	5.4%
Vietnam	7.9%	6.3%
<i>Slow Growth Club (2% or less per year in both time periods)</i>		
Central African Republic	2.0%	−0.2%
France	2.0%	1.4%
Germany	1.8%	2.0%
Haiti	−1.5%	1.5%
Italy	1.6%	0.3%

TABLE 20.5 Economic Growth around the World (Source: http://databank.worldbank.org/data/views/variableSelection/selectvariables.aspx?source=world-development-indicators#c_u)

Country	Average Growth Rate of Real GDP 1990–2000	Growth Rate of Real GDP 2010–2019
Jamaica	0.9%	0.7%
Japan	1.3%	1.3%
Switzerland	1.0%	2.0%
United States (for reference)	3.2%	2.3%
<i>World Overview</i>		
High income	2.7%	1.7%
Low income	3.8%	4.5%
Middle income	4.7%	4.0%

TABLE 20.5 Economic Growth around the World (Source: http://databank.worldbank.org/data/views/variableSelection/selectvariables.aspx?source=world-development-indicators#c_u)

Each of the countries in [Table 20.5](#) has its own unique story of investments in human and physical capital, technological gains, market forces, government policies, and even lucky events, but an overall pattern of convergence is clear. The low-income countries have GDP growth that is faster than that of the middle-income countries, which in turn have GDP growth that is faster than that of the high-income countries. Two prominent members of the fast-growth club are China and India, which between them have nearly 40% of the world's population. Some prominent members of the slow-growth club are high-income countries like France, Germany, Italy, and Japan.

Will this pattern of economic convergence persist into the future? This is a controversial question among economists that we will consider by looking at some of the main arguments on both sides.

Arguments Favoring Convergence

Several arguments suggest that low-income countries might have an advantage in achieving greater worker productivity and economic growth in the future.

A first argument is based on diminishing marginal returns. Even though deepening human and physical capital will tend to increase GDP per capita, the law of diminishing returns suggests that as an economy continues to increase its human and physical capital, the marginal gains to economic growth will diminish. For example, raising the average education level of the population by two years from a tenth-grade level to a high school diploma (while holding all other inputs constant) would produce a certain increase in output. An additional two-year increase, so that the average person had a two-year college degree, would increase output further, but the marginal gain would be smaller. Yet another additional two-year increase in the level of education, so that the average person would have a four-year-college bachelor's degree, would increase output still further, but the marginal increase would again be smaller. A similar lesson holds for physical capital. If the quantity of physical capital available to the average worker increases, by, say, \$5,000 to \$10,000 (again, while holding all other inputs constant), it will increase the level of output. An additional increase from \$10,000 to \$15,000 will increase output further, but the marginal increase will be smaller.

Low-income countries like China and India tend to have lower levels of human capital and physical capital, so an investment in capital deepening should have a larger marginal effect in these countries than in high-

income countries, where levels of human and physical capital are already relatively high. Diminishing returns implies that low-income economies could converge to the levels that the high-income countries achieve.

A second argument is that low-income countries may find it easier to improve their technologies than high-income countries. High-income countries must continually invent new technologies, whereas low-income countries can often find ways of applying technology that has already been invented and is well understood. The economist Alexander Gerschenkron (1904–1978) gave this phenomenon a memorable name: “the advantages of backwardness.” Of course, he did not literally mean that it is an advantage to have a lower standard of living. He was pointing out that a country that is behind has some extra potential for catching up.

Finally, optimists argue that many countries have observed the experience of those that have grown more quickly and have learned from it. Moreover, once the people of a country begin to enjoy the benefits of a higher standard of living, they may be more likely to build and support the market-friendly institutions that will help provide this standard of living.

LINK IT UP

View this [video \(http://openstax.org/l/tedhansrosling\)](http://openstax.org/l/tedhansrosling) to learn about economic growth across the world.

Arguments That Convergence Is neither Inevitable nor Likely

If the economy's growth depended only on the deepening of human capital and physical capital, then we would expect that economy's growth rate to slow down over the long run because of diminishing marginal returns. However, there is another crucial factor in the aggregate production function: technology.

Developing new technology can provide a way for an economy to sidestep the diminishing marginal returns of capital deepening. [Figure 20.7](#) shows how. The figure's horizontal axis measures the amount of capital deepening, which on this figure is an overall measure that includes deepening of both physical and human capital. The amount of human and physical capital per worker increases as you move from left to right, from C_1 to C_2 to C_3 . The diagram's vertical axis measures per capita output. Start by considering the lowest line in this diagram, labeled Technology 1. Along this aggregate production function, the level of technology is held constant, so the line shows only the relationship between capital deepening and output. As capital deepens from C_1 to C_2 to C_3 and the economy moves from R to U to W, per capita output does increase—but the way in which the line starts out steeper on the left but then flattens as it moves to the right shows the diminishing marginal returns, as additional marginal amounts of capital deepening increase output by ever-smaller amounts. The shape of the aggregate production line (Technology 1) shows that the ability of capital deepening, by itself, to generate sustained economic growth is limited, since diminishing returns will eventually set in.

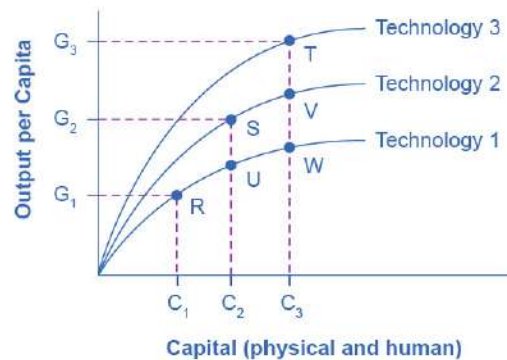


FIGURE 20.7 Capital Deepening and New Technology Imagine that the economy starts at point R, with the level of physical and human capital C_1 and the output per capita at G_1 . If the economy relies only on capital deepening, while remaining at the technology level shown by the Technology 1 line, then it would face diminishing marginal

returns as it moved from point R to point U to point W. However, now imagine that capital deepening is combined with improvements in technology. Then, as capital deepens from C_1 to C_2 , technology improves from Technology 1 to Technology 2, and the economy moves from R to S. Similarly, as capital deepens from C_2 to C_3 , technology increases from Technology 2 to Technology 3, and the economy moves from S to T. With improvements in technology, there is no longer any reason that economic growth must necessarily slow down.

Now, bring improvements in technology into the picture. Improved technology means that with a given set of inputs, more output is possible. The production function labeled Technology 1 in the figure is based on one level of technology, but Technology 2 is based on an improved level of technology, so for every level of capital deepening on the horizontal axis, it produces a higher level of output on the vertical axis. In turn, production function Technology 3 represents a still higher level of technology, so that for every level of inputs on the horizontal axis, it produces a higher level of output on the vertical axis than either of the other two aggregate production functions.

Most healthy, growing economies are deepening their human and physical capital and increasing technology at the same time. As a result, the economy can move from a choice like point R on the Technology 1 aggregate production line to a point like S on Technology 2 and a point like T on the still higher aggregate production line (Technology 3). With the combination of technology and capital deepening, the rise in GDP per capita in high-income countries does not need to fade away because of diminishing returns. The gains from technology can offset the diminishing returns involved with capital deepening.

Will technological improvements themselves run into diminishing returns over time? That is, will it become continually harder and more costly to discover new technological improvements? Perhaps someday, but, at least over the last two centuries since the beginning of the Industrial Revolution, improvements in technology have not run into diminishing marginal returns. Modern inventions, like the internet or discoveries in genetics or materials science, do not seem to provide smaller gains to output than earlier inventions like the steam engine or the railroad. One reason that technological ideas do not seem to run into diminishing returns is that we often can apply widely the ideas of new technology at a marginal cost that is very low or even zero. A specific worker or group of workers must use a specific additional machine, or an additional year of education. Many workers across the economy can use a new technology or invention at very low marginal cost.

The argument that it is easier for a low-income country to copy and adapt existing technology than it is for a high-income country to invent new technology is not necessarily true, either. When it comes to adapting and using new technology, a society's performance is not necessarily guaranteed, but is the result of whether the country's economic, educational, and public policy institutions are supportive. In theory, perhaps, low-income countries have many opportunities to copy and adapt technology, but if they lack the appropriate supportive economic infrastructure and institutions, the theoretical possibility that backwardness might have certain advantages is of little practical relevance.

LINK IT UP

Visit this [website \(http://openstax.org/l/Indiapoverty\)](http://openstax.org/l/Indiapoverty) to read more about economic growth in India.

The Slowness of Convergence

Although economic convergence between the high-income countries and the rest of the world seems possible and even likely, it will proceed slowly. Consider, for example, a country that starts off with a GDP per capita of \$40,000, which would roughly represent a typical high-income country today, and another country that starts out at \$4,000, which is roughly the level in low-income but not impoverished countries like Indonesia, Guatemala, or Egypt. Say that the rich country chugs along at a 2% annual growth rate of GDP per capita, while the poorer country grows at the aggressive rate of 7% per year. After 30 years, GDP per capita in the rich country will be \$72,450 (that is, $\$40,000 (1 + 0.02)^{30}$) while in the poor country it will be \$30,450 (that is, $\$4,000 (1 + 0.07)^{30}$). Convergence has occurred. The rich country used to be 10 times as wealthy as the poor

one, and now it is only about 2.4 times as wealthy. Even after 30 consecutive years of very rapid growth, however, people in the low-income country are still likely to feel quite poor compared to people in the rich country. Moreover, as the poor country catches up, its opportunities for catch-up growth are reduced, and its growth rate may slow down somewhat.

The slowness of convergence illustrates again that small differences in annual rates of economic growth become huge differences over time. The high-income countries have been building up their advantage in standard of living over decades—more than a century in some cases. Even in an optimistic scenario, it will take decades for the low-income countries of the world to catch up significantly.



BRING IT HOME

Calories and Economic Growth

We can tell the story of modern economic growth by looking at calorie consumption over time. The dramatic rise in incomes allowed the average person to eat better and consume more calories. How did these incomes increase? The neoclassical growth consensus uses the aggregate production function to suggest that the period of modern economic growth came about because of increases in inputs such as technology and physical and human capital. Also important was the way in which technological progress combined with physical and human capital deepening to create growth and convergence. The issue of distribution of income notwithstanding, it is clear that the average worker can afford more calories in 2020 than in 1875.

Aside from increases in income, there is another reason why the average person can afford more food. Modern agriculture has allowed many countries to produce more food than they need. Despite having more than enough food, however, many governments and multilateral agencies have not solved the food distribution problem. In fact, food shortages, famine, or general food insecurity are caused more often by the failure of government macroeconomic policy, according to the Nobel Prize-winning economist Amartya Sen. Sen has conducted extensive research into issues of inequality, poverty, and the role of government in improving standards of living. Macroeconomic policies that strive toward stable inflation, full employment, education of women, and preservation of property rights are more likely to eliminate starvation and provide for a more even distribution of food.

Because we have more food per capita, global food prices have decreased since 1875. The prices of some foods, however, have decreased more than the prices of others. For example, researchers from the University of Washington have shown that in the United States, calories from zucchini and lettuce are 100 times more expensive than calories from oil, butter, and sugar. Research from countries like India, China, and the United States suggests that as incomes rise, individuals want more calories from fats and protein and fewer from carbohydrates. This has very interesting implications for global food production, obesity, and environmental consequences. Affluent urban India has an obesity problem much like many parts of the United States. The forces of convergence are at work.

Key Terms

aggregate production function the process whereby an economy as a whole turns economic inputs such as human capital, physical capital, and technology into output measured as GDP per capita

capital deepening an increase by society in the average level of physical and/or human capital per person

compound growth rate the rate of growth when multiplied by a base that includes past GDP growth

contractual rights the rights of individuals to enter into agreements with others regarding the use of their property providing recourse through the legal system in the event of noncompliance

convergence pattern in which economies with low per capita incomes grow faster than economies with high per capita incomes

human capital the accumulated skills and education of workers

Industrial Revolution the widespread use of power-driven machinery and the economic and social changes that occurred in the first half of the 1800s

infrastructure a component of physical capital such as roads and rail systems

innovation putting advances in knowledge to use in a new product or service

invention advances in knowledge

labor productivity the value of what is produced per worker, or per hour worked (sometimes called worker productivity)

modern economic growth the period of rapid economic growth from 1870 onward

physical capital the plant and equipment that firms use in production; this includes infrastructure

production function the process whereby a firm turns economic inputs like labor, machinery, and raw materials into outputs like goods and services that consumers use

rule of law the process of enacting laws that protect individual and entity rights to use their property as they see fit. Laws must be clear, public, fair, and enforced, and applicable to all members of society

special economic zone (SEZ) area of a country, usually with access to a port where, among other benefits, the government does not tax trade

technological change a combination of invention—advances in knowledge—and innovation

technology all the ways in which existing inputs produce more or higher quality, as well as different and altogether new products

Key Concepts and Summary

20.1 The Relatively Recent Arrival of Economic Growth

Since the early nineteenth century, there has been a spectacular process of long-run economic growth during which the world's leading economies—mostly those in Western Europe and North America—expanded GDP per capita at an average rate of about 2% per year. In the last half-century, countries like Japan, South Korea, and China have shown the potential to catch up. The Industrial Revolution facilitated the extensive process of economic growth, that economists often refer to as modern economic growth. This increased worker productivity and trade, as well as the development of governance and market institutions.

20.2 Labor Productivity and Economic Growth

We can measure productivity, the value of what is produced per worker, or per hour worked, as the level of GDP per worker or GDP per hour. The United States experienced a productivity slowdown between 1973 and 1989. Since then, U.S. productivity has rebounded for the most part, but annual growth in productivity in the nonfarm business sector has been less than one percent each year between 2011 and 2016. It is not clear what productivity growth will be in the coming years. The rate of productivity growth is the primary determinant of an economy's rate of long-term economic growth and higher wages. Over decades and generations, seemingly small differences of a few percentage points in the annual rate of economic growth make an enormous difference in GDP per capita. An aggregate production function specifies how certain inputs in the economy, like human capital, physical capital, and technology, lead to the output measured as GDP per capita.

Compound interest and compound growth rates behave in the same way as productivity rates. Seemingly small changes in percentage points can have big impacts on income over time.

20.3 Components of Economic Growth

Over decades and generations, seemingly small differences of a few percentage points in the annual rate of economic growth make an enormous difference in GDP per capita. Capital deepening refers to an increase in the amount of capital per worker, either human capital per worker, in the form of higher education or skills, or physical capital per worker. Technology, in its economic meaning, refers broadly to all new methods of production, which includes major scientific inventions but also small inventions and even better forms of management or other types of institutions. A healthy climate for growth in GDP per capita consists of improvements in human capital, physical capital, and technology, in a market-oriented environment with supportive public policies and institutions.

20.4 Economic Convergence

When countries with lower GDP levels per capita catch up to countries with higher GDP levels per capita, we call the process convergence. Convergence can occur even when both high- and low-income countries increase investment in physical and human capital with the objective of growing GDP. This is because the impact of new investment in physical and human capital on a low-income country may result in huge gains as new skills or equipment combine with the labor force. In higher-income countries, however, a level of investment equal to that of the low income country is not likely to have as big an impact, because the more developed country most likely already has high levels of capital investment. Therefore, the marginal gain from this additional investment tends to be successively less and less. Higher income countries are more likely to have diminishing returns to their investments and must continually invent new technologies. This allows lower-income economies to have a chance for convergent growth. However, many high-income economies have developed economic and political institutions that provide a healthy economic climate for an ongoing stream of technological innovations. Continuous technological innovation can counterbalance diminishing returns to investments in human and physical capital.

Self-Check Questions

1. Explain what the Industrial Revolution was and where it began.
2. Explain the difference between property rights and contractual rights. Why do they matter to economic growth?
3. Are there other ways in which we can measure productivity besides the amount produced per hour of work?
4. Assume there are two countries: South Korea and the United States. South Korea grows at 4% and the United States grows at 1%. For the sake of simplicity, assume they both start from the same fictional income level, \$10,000. What will the incomes of the United States and South Korea be in 20 years? By how many multiples will each country's income grow in 20 years?
5. What do the growth accounting studies conclude are the determinants of growth? Which is more important, the determinants or how they are combined?
6. What policies can the government of a free-market economy implement to stimulate economic growth?
7. List the areas where government policy can help economic growth.
8. Use an example to explain why, after periods of rapid growth, a low-income country that has not caught up to a high-income country may feel poor.

9. Would the following events usually lead to capital deepening? Why or why not?
 - a. A weak economy in which businesses become reluctant to make long-term investments in physical capital.
 - b. A rise in international trade.
 - c. A trend in which many more adults participate in continuing education courses through their employers and at colleges and universities.
10. What are the “advantages of backwardness” for economic growth?
11. Would you expect capital deepening to result in diminished returns? Why or why not? Would you expect improvements in technology to result in diminished returns? Why or why not?
12. Why does productivity growth in high-income economies not slow down as it runs into diminishing returns from additional investments in physical capital and human capital? Does this show one area where the theory of diminishing returns fails to apply? Why or why not?

Review Questions

13. How did the Industrial Revolution increase the economic growth rate and income levels in the United States?
14. How much should a nation be concerned if its rate of economic growth is just 2% slower than other nations?
15. How is GDP per capita calculated differently from labor productivity?
16. How do gains in labor productivity lead to gains in GDP per capita?
17. What is an aggregate production function?
18. What is capital deepening?
19. What do economists mean when they refer to improvements in technology?
20. For a high-income economy like the United States, what aggregate production function elements are most important in bringing about growth in GDP per capita? What about a middle-income country such as Brazil? A low-income country such as Niger?
21. List some arguments for and against the likelihood of convergence.

Critical Thinking Questions

22. Over the past 50 years, many countries have experienced an annual growth rate in real GDP per capita greater than that of the United States. Some examples are China, Japan, South Korea, and Taiwan. Does that mean the United States is regressing relative to other countries? Does that mean these countries will eventually overtake the United States in terms of the growth rate of real GDP per capita? Explain.
23. [Labor Productivity and Economic Growth](#) outlined the logic of how increased productivity is associated with increased wages. Detail a situation where this is not the case and explain why it is not.
24. Change in labor productivity is one of the most watched international statistics of growth. Visit the St. Louis Federal Reserve website and find the data section (<http://research.stlouisfed.org>). Find international comparisons of labor productivity, listed under the FRED Economic database (Growth Rate of Total Labor Productivity), and compare two countries in the recent past. State what you think the reasons for differences in labor productivity could be.
25. Refer back to the [Work It Out](#) about Comparing the Economies of Two Countries and examine the data for the two countries you chose. How are they similar? How are they different?

26. Education seems to be important for human capital deepening. As people become better educated and more knowledgeable, are there limits to how much additional benefit more education can provide? Why or why not?
27. Describe some of the political and social tradeoffs that might occur when a less developed country adopts a strategy to promote labor force participation and economic growth via investment in girls' education.
28. Why is investing in girls' education beneficial for growth?
29. How is the concept of technology, as defined with the aggregate production function, different from our everyday use of the word?
30. What sorts of policies can governments implement to encourage convergence?
31. As technological change makes us more sedentary and food costs increase, obesity is likely. What factors do you think may limit obesity?

Problems

32. An economy starts off with a GDP per capita of \$5,000. How large will the GDP per capita be if it grows at an annual rate of 2% for 20 years? 2% for 40 years? 4% for 40 years? 6% for 40 years?
33. An economy starts off with a GDP per capita of 12,000 euros. How large will the GDP per capita be if it grows at an annual rate of 3% for 10 years? 3% for 30 years? 6% for 30 years?
34. Say that the average worker in Canada has a productivity level of \$30 per hour while the average worker in the United Kingdom has a productivity level of \$25 per hour (both measured in U.S. dollars). Over the next five years, say that worker productivity in Canada grows at 1% per year while worker productivity in the UK grows 3% per year. After five years, who will have the higher productivity level, and by how much?
35. Say that the average worker in the U.S. economy is eight times as productive as an average worker in Mexico. If the productivity of U.S. workers grows at 2% for 25 years and the productivity of Mexico's workers grows at 6% for 25 years, which country will have higher worker productivity at that point?



FIGURE 21.1 Out of Business Borders was one of the many companies unable to recover from the 2008-2009 economic recession. (Credit: modification of "Factory Automation Robotics Palettizing Bread" by KUKA Roboter GmbH/Wikimedia Commons, Public Domain)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- How Economists Define and Compute Unemployment Rate
- Patterns of Unemployment
- What Causes Changes in Unemployment over the Short Run
- What Causes Changes in Unemployment over the Long Run

Introduction to Unemployment



BRING IT HOME

Unemployment and the COVID-19 Pandemic: A Complicated Story

It was the most abrupt economic change in the post-World War II era. Between March 2020 and April 2020, the U.S. unemployment rate increased from 4.4% to 14.8%. As a result of the COVID-19 pandemic, millions of people were left without work as businesses shut down and people stayed home and cut their spending, especially on restaurants, tourism, and travel. As confidence and spending were slowly restored, and as the situation with the virus steadily improved, unemployment began to tick down. By 2022, with the availability of vaccines and boosters and other improved health measures, things were better still, but the presence of dangerous variants prevented a full return to normal.

The COVID-19 pandemic had other effects on the labor market as well. Labor force participation—the rate at which people are employed or actively searching for work—declined and as of early-2022 remained lower than it was in 2019. Some were forced to stop working because of school and childcare closures. Others were concerned about how safe their workplaces would be in the middle of a global pandemic. And still others simply chose to retire early. Labor force participation remains a sore spot in the labor market's recovery.

These two statistics—unemployment and labor force participation—show how complicated the labor market can be. As the unemployment rate declined through 2021, the disappointing statistics on labor force participation show weak points. One day you may have read a headline about how easy it was to find a job, and the next day a headline would describe how difficult it was for employers to find workers. By the end of this chapter, you will be in a much better position to make sense of these events.

Unemployment can be a terrible and wrenching life experience—like a serious automobile accident or a messy divorce—whose consequences only someone who has gone through it can fully understand. For unemployed individuals and their families, there is the day-to-day financial stress of not knowing from where the next paycheck is coming. There are painful adjustments, like watching your savings account dwindle, selling a car and buying a cheaper one, or moving to a less expensive place to live. Even when the unemployed person finds a new job, it may pay less than the previous one. For many people, their job is an important part of their self worth. When unemployment separates people from the workforce, it can affect family relationships as well as mental and physical health.

The human costs of unemployment alone would justify making a low level of unemployment an important public policy priority. However, unemployment also includes economic costs to the broader society. When millions of unemployed but willing workers cannot find jobs, economic resource are unused. An economy with high unemployment is like a company operating with a functional but unused factory. The opportunity cost of unemployment is the output that the unemployed workers could have produced.

This chapter will discuss how economists define and compute the unemployment rate. It will examine the patterns of unemployment over time, for the U.S. economy as a whole, for different demographic groups in the U.S. economy, and for other countries. It will then consider an economic explanation for unemployment, and how it explains the patterns of unemployment and suggests public policies for reducing it.

21.1 How Economists Define and Compute Unemployment Rate

LEARNING OBJECTIVES

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

- Calculate the labor force participation rate and the unemployment rate
- Explain hidden unemployment and what it means to be in or out of the labor force
- Evaluate the collection and interpretation of unemployment data

Newspaper or television reports typically describe unemployment as a percentage or a rate. A recent report might have said, for example, *from September 2021 to October 2021, the U.S. unemployment rate declined from 4.8% to 4.6%*. At a glance, the changes between the percentages may seem small. However, remember that the U.S. economy has about 162 million adults (as of the beginning of 2022) who either have jobs or are looking for them. A rise or fall of just 0.1% in the unemployment rate of 162 million potential workers translates into 160,000 people, which is roughly the total population of a city like Syracuse, New York, Brownsville, Texas, or Pasadena, California. Large rises in the unemployment rate mean large numbers of job losses. In April 2020, at the peak of the pandemic-induced recession, over 20 million people were out of work. Even with the unemployment rate at 4.2% in November 2021, about 7 million people who were looking for jobs were out of work.

LINK IT UP

The [Bureau of Labor Statistics \(http://openstax.org/l/BLS1\)](http://openstax.org/l/BLS1) tracks and reports all data related to unemployment.

Who's In or Out of the Labor Force?

Should we count everyone without a job as unemployed? Of course not. For example, we should not count children as unemployed. Surely, we should not count the retired as unemployed. Many full-time college students have only a part-time job, or no job at all, but it seems inappropriate to count them as suffering the pains of unemployment. Some people are not working because they are rearing children, ill, on vacation, or on parental leave.

The point is that we do not just divide the adult population into employed and unemployed. A third group exists: people who do not have a job, and for some reason—retirement, looking after children, taking a voluntary break before a new job—are not interested in having a job, either. It also includes those who do want a job but have quit looking, often due to discouragement due to their inability to find suitable employment. Economists refer to this third group of those who are not working and not looking for work as **out of the labor force** or not in the labor force.

The U.S. unemployment rate, which is based on a monthly survey carried out by the U.S. Bureau of the Census, asks a series of questions to divide the adult population into employed, unemployed, or not in the labor force. To be classified as unemployed, a person must be without a job, currently available to work, and actively looking for work in the previous four weeks. Thus, a person who does not have a job but who is not currently available to work or has not actively looked for work in the last four weeks is counted as out of the labor force.

Employed: currently working for pay

Unemployed: Out of work and actively looking for a job

Out of the labor force: Out of paid work and not actively looking for a job

Labor force: the number of employed plus the unemployed

Calculating the Unemployment Rate

[Figure 21.2](#) shows the three-way division of the 16-and-over population. In January 2017, about 62.9% of the adult population was "in the labor force"; that is, people are either employed or without a job but looking for work. We can divide those in the labor force into the employed and the unemployed. [Table 21.1](#) shows those values. The **unemployment rate** is not the percentage of the total adult population without jobs, but rather the percentage of adults who are in the labor force but who do not have jobs:

$$\text{Unemployment rate} = \frac{\text{Unemployed people}}{\text{Total labor force}} \times 100$$

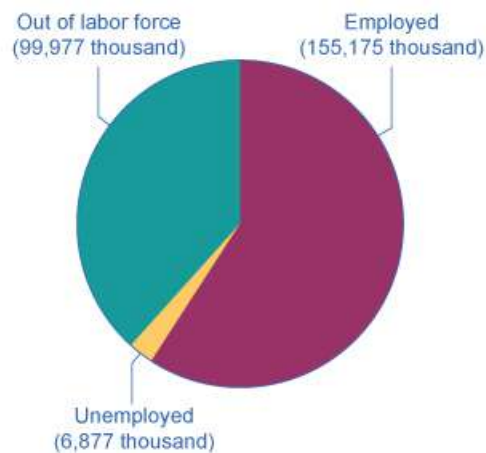


FIGURE 21.2 Employed, Unemployed, and Out of the Labor Force Distribution of Adult Population (age 16 and older), January 2021 The total adult, working-age population in January 2021 was 262.029 million. Out of this total population, 155.175 million were classified as employed, and 6.877 million were classified as unemployed. The remaining 99.977 million were classified as out of the labor force. As you will learn, however, this seemingly simple chart does not tell the whole story. As you will learn, however, this seemingly simple chart does not tell the whole story.

Total adult population over the age of 16	262.029 million
In the labor force	162.052 million (59.2%)
Employed	155.175 million
Unemployed	6.877 million
Out of the labor force	99.977 million (38.2%)

TABLE 21.1 U.S. Employment and Unemployment, November 2021
(Source: <https://data.bls.gov>)

In this example, we can calculate the unemployment rate as 7.635 million unemployed people divided by 159.716 million people in the labor force, which works out to a 4.8% rate of unemployment. The following Work It Out feature will walk you through the steps of this calculation.

WORK IT OUT

Calculating Labor Force Percentages

How do economists arrive at the percentages in and out of the labor force and the unemployment rate? We will use the values in [Table 21.1](#) to illustrate the steps.

To determine the percentage in the labor force:

Step 1. Divide the number of people in the labor force (159.716 million) by the total adult (working-age) population (254.082 million).

Step 2. Multiply by 100 to obtain the percentage.

$$\begin{aligned}\text{Percentage out of the labor force} &= \frac{159.716}{254.082} \\ &= 0.6286 \\ &= 62.9\%\end{aligned}$$

To determine the percentage out of the labor force:

Step 1. Divide the number of people out the labor force (94.366 million) by the total adult (working-age) population (254.082 million).

Step 2. Multiply by 100 to obtain the percentage.

$$\begin{aligned}\text{Percentage in the labor force} &= \frac{94.366}{254.082} \\ &= 0.3714 \\ &= 37.1\%\end{aligned}$$

To determine the unemployment rate:

Step 1. Divide the number of unemployed people (7.635 million) by the total labor force (157 million).

Step 2. Multiply by 100 to obtain the rate.

$$\begin{aligned}\text{Unemployment rate} &= \frac{7.635}{159.716} \\ &= 0.0478 \\ &= 4.8\%\end{aligned}$$

Hidden Unemployment

Even with the “out of the labor force” category, there are still some people who are mislabeled in the categorization of employed, unemployed, or out of the labor force. There are some people who have only part time or temporary jobs, and they are looking for full time and permanent employment that are counted as employed, although they are not employed in the way they would like or need to be. Additionally, there are individuals who are **underemployed**. This includes those who are trained or skilled for one type or level of work but are working in a lower paying job or one that does not utilize their skills. For example, we would consider an individual with a college degree in finance who is working as a sales clerk underemployed. They are, however, also counted in the employed group. All of these individuals fall under the umbrella of the term “hidden unemployment.” **Discouraged workers**, those who have stopped looking for employment and, hence, are no longer counted in the unemployed also fall into this group

Labor Force Participation Rate

Another important statistic is the **labor force participation rate**. This is the percentage of adults in an economy who are either employed or who are unemployed and looking for a job. Using the data in [Figure 21.2](#) and [Table 21.1](#), those included in this calculation would be the 162.052 million individuals in the labor force. We calculate the rate by taking the number of people in the labor force, that is, the number employed and the number unemployed, divided by the total adult population and multiplying by 100 to get the percentage. For the data from November 2021, the labor force participation rate is 61.8%. Historically, the civilian labor force participation rate in the United States climbed beginning in the 1960s as women increasingly entered the workforce, and it peaked at just over 67% in late 1999 to early 2000. Since then, the labor force participation rate has steadily declined, slowly to about 66% in 2008, early in the Great Recession, and then more rapidly during and after that recession. The labor force then climbed slowly during the 2010s but declined again during the pandemic in March–April 2020 and remained lower than pre-pandemic levels as of early 2022.

The Establishment Payroll Survey

When the unemployment report comes out each month, the Bureau of Labor Statistics (BLS) also reports on the number of jobs created—which comes from the establishment payroll survey. The payroll survey is based on a survey of about 147,000 businesses and government agencies throughout the United States. It generates payroll employment estimates by the following criteria: all employees, average weekly hours worked, and average hourly, weekly, and overtime earnings. One of the criticisms of this survey is that it does not count the self-employed. It also does not make a distinction between new, minimum wage, part time or temporary jobs and full time jobs with “decent” pay.

How Does the U.S. Bureau of Labor Statistics Collect the U.S. Unemployment Data?

The unemployment rate announced by the U.S. Bureau of Labor Statistics on the first Friday of each month for the previous month is based on the Current Population Survey (CPS), which the Bureau has carried out every month since 1940. The Bureau takes great care to make this survey representative of the country as a whole. The country is first divided into 3,137 areas. The U.S. Bureau of the Census then selects 729 of these areas to survey. It divides the 729 areas into districts of about 300 households each, and divides each district into clusters of about four dwelling units. Every month, Census Bureau employees call about 15,000 of the four-household clusters, for a total of 60,000 households. Employees interview households for four consecutive months, then rotate them out of the survey for eight months, and then interview them again for the same four months the following year, before leaving the sample permanently.

Based on this survey, state, industry, urban and rural areas, gender, age, race or ethnicity, and level of education statistics comprise components that contribute to unemployment rates. A wide variety of other information is available, too. For example, how long have people been unemployed? Did they become unemployed because they quit, or were laid off, or their employer went out of business? Is the unemployed person the only wage earner in the family? The Current Population Survey is a treasure trove of information about employment and unemployment. If you are wondering what the difference is between the CPS and EPS, read the following Clear it Up feature.



CLEAR IT UP

What is the difference between CPS and EPS?

The United States Census Bureau conducts the Current Population Survey (CPS), which measures the percentage of the labor force that is unemployed. The Bureau of Labor Statistics' establishment payroll survey (EPS) is a payroll survey that measures the net change in jobs created for the month.

Criticisms of Measuring Unemployment

There are always complications in measuring the number of unemployed. For example, what about people who do not have jobs and would be available to work, but are discouraged by the lack of available jobs in their area and stopped looking? Such people, and their families, may be suffering the pains of unemployment. However, the survey counts them as out of the labor force because they are not actively looking for work. Other people may tell the Census Bureau that they are ready to work and looking for a job but, truly, they are not that eager to work and are not looking very hard at all. They are counted as unemployed, although they might more accurately be classified as out of the labor force. Still other people may have a job, perhaps doing something like yard work, child care, or cleaning houses, but are not reporting the income earned to the tax authorities. They may report being unemployed, when they actually are working.

Although the unemployment rate gets most of the public and media attention, economic researchers at the Bureau of Labor Statistics publish a wide array of surveys and reports that try to measure these kinds of issues and to develop a more nuanced and complete view of the labor market. It is not exactly a hot news flash that

economic statistics are imperfect. Even imperfect measures like the unemployment rate, however, can still be quite informative, when interpreted knowledgeably and sensibly.

LINK IT UP

Click [here \(http://openstax.org/l/BLS_CPS\)](http://openstax.org/l/BLS_CPS) to learn more about the CPS and to read frequently asked questions about employment and labor.

21.2 Patterns of Unemployment

LEARNING OBJECTIVES

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

- Explain historical patterns of unemployment in the U.S.
- Identify trends of unemployment based on demographics
- Evaluate global unemployment rates

Let's look at how unemployment rates have changed over time and how various groups of people are affected by unemployment differently.

The Historical U.S. Unemployment Rate

[Figure 21.3](#) shows the historical pattern of U.S. unemployment since 1955.

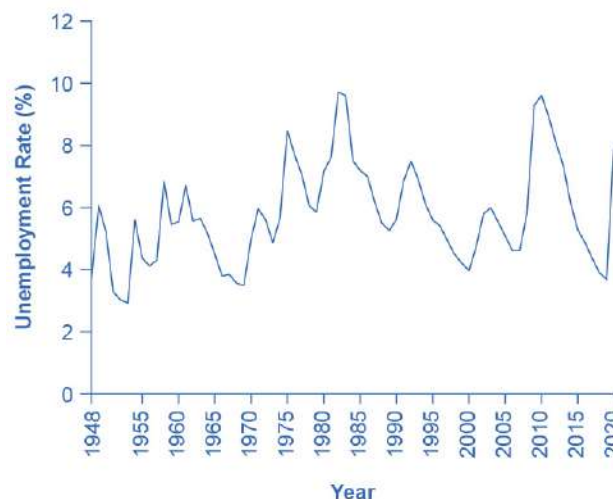


FIGURE 21.3 The U.S. Unemployment Rate, 1948–2020 The U.S. unemployment rate moves up and down as the economy moves in and out of recessions. However, over time, the unemployment rate seems to return to a range of 4% to 6%. There does not seem to be a long-term trend toward the rate moving generally higher or generally lower. (Source: Federal Reserve Economic Data (FRED) <https://research.stlouisfed.org/fred2/series/LRUN64TTUSA156S0>)

As we look at this data, several patterns stand out:

1. Unemployment rates do fluctuate over time. During the deep recessions of the early 1980s, 2007–2009, and 2021, unemployment reached roughly 10%. For comparison, during the 1930s Great Depression, the unemployment rate reached almost 25% of the labor force.
2. Unemployment rates in the late 1990s and into the mid-2000s were rather low by historical standards. The unemployment rate was below 5% from 1997 to 2000, and near 5% during almost all of 2006–2007, and 5% or less from September 2015 through March 2020. The previous time unemployment had been less than 5% for three consecutive years was three decades earlier, from 1968 to 1970.
3. The unemployment rate never falls all the way to zero. It almost never seems to get below 3%—and it stays that low only for very short periods. (We discuss reasons why this is the case later in this chapter.)
4. The timing of rises and falls in unemployment matches fairly well with the timing of upswings and

downswings in the overall economy, except that unemployment tends to lag changes in economic activity, and especially so during upswings of the economy following a recession. During periods of recession and depression, unemployment is high. During periods of economic growth, unemployment tends to be lower.

5. No significant upward or downward trend in unemployment rates is apparent. This point is especially worth noting because the U.S. population more than quadrupled from 76 million in 1900 to over 324 million by 2017. Moreover, a higher proportion of U.S. adults are now in the paid workforce, because women have entered the paid labor force in significant numbers in recent decades. Women comprised 18% of the paid workforce in 1900 and nearly half of the paid workforce in 2021. However, despite the increased number of workers, as well as other economic events like globalization and the continuous invention of new technologies, the economy has provided jobs without causing any long-term upward or downward trend in unemployment rates.

[Click to view content \(https://openstax.org/books/principles-economics-3e/pages/21-2-patterns-of-unemployment\)](https://openstax.org/books/principles-economics-3e/pages/21-2-patterns-of-unemployment)

Unemployment Rate from 1948. Similar to the graph in the figure above, the rate moves up and down as the economy moves in and out of recessions.

Unemployment Rates by Group

Unemployment is not distributed evenly across the U.S. population. [Figure 21.4](#) shows unemployment rates broken down in various ways: by gender, age, and race/ethnicity.

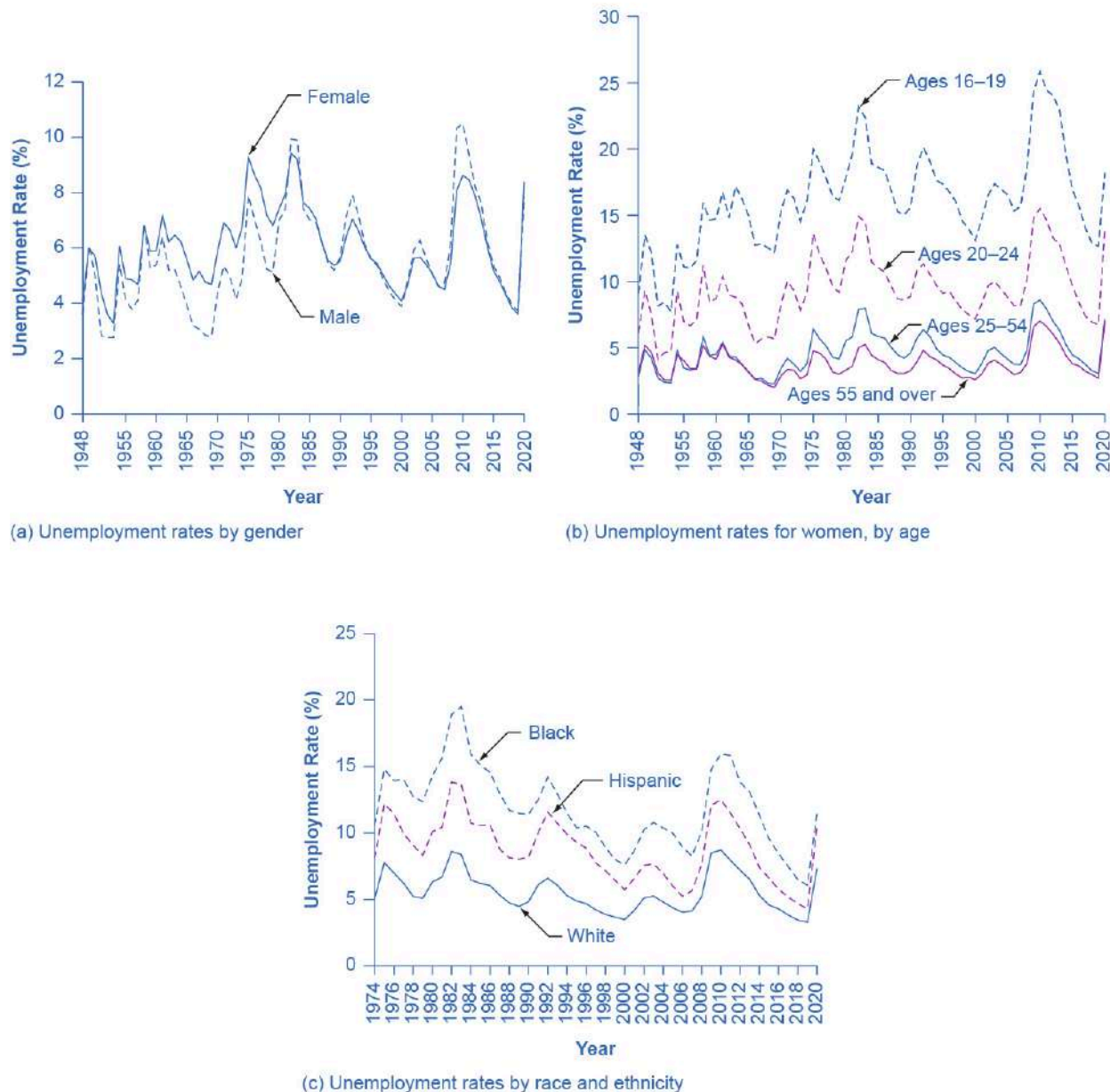


FIGURE 21.4 Unemployment Rate by Demographic Group (a) By gender, 1948–2020. Unemployment rates for men used to be lower than unemployment rates for women, but in recent decades, the two rates have been very close, often— and especially during and soon after the Great Recession – with the unemployment rate for men somewhat higher. (b) By age, 1948–2020. Unemployment rates are highest for the very young and become lower with age. (c) By race and ethnicity, 1974–2020. Although unemployment rates for all groups tend to rise and fall together, the unemployment rate for Black people is typically about twice as high as that for White people, while the unemployment rate for Hispanic people is in between. (Source: www.bls.gov)

The unemployment rate for women had historically tended to be higher than the unemployment rate for men, perhaps reflecting the historical pattern that women were seen as “secondary” earners. By about 1980, however, the unemployment rate for women was essentially the same as that for men, as [Figure 21.4](#) (a) shows. During the pandemic-induced recession of 2020 and in the immediate aftermath, the unemployment rate for women exceeded the unemployment rate for men. Subsequently, however, the gap has narrowed.

LINK IT UP

Read this [report \(http://openstax.org/l/BLS_recession\)](http://openstax.org/l/BLS_recession) for detailed information on the 2008–2009 recession. It also provides some very useful information on the statistics of unemployment.

Younger workers tend to have higher unemployment, while middle-aged workers tend to have lower unemployment, probably because the middle-aged workers feel the responsibility of needing to have a job more heavily. Younger workers move in and out of jobs more than middle-aged workers, as part of the process of matching of workers and jobs, and this contributes to their higher unemployment rates. In addition, middle-aged workers are more likely to feel the responsibility of needing to have a job more heavily. Elderly workers have extremely low rates of unemployment, because those who do not have jobs often exit the labor force by retiring, and thus are not counted in the unemployment statistics. [Figure 21.4](#) (b) shows unemployment rates for women divided by age. The pattern for men is similar.

The unemployment rate for African-Americans is substantially higher than the rate for other racial or ethnic groups, a fact that surely reflects, to some extent, a pattern of discrimination that has constrained Black people's labor market opportunities. However, the gaps between unemployment rates for White, Black, and Hispanic people have diminished in the 1990s, as [Figure 21.4](#) (c) shows. In fact, unemployment rates for Black and Hispanic people were at the lowest levels for several decades in the mid-2000s before rising during the recent Great Recession.

Finally, those with less education typically suffer higher unemployment. In November 2021, for example, the unemployment rate for those with a college degree was 2.3%; for those with some college but not a four year degree, the unemployment rate was 3.7%; for high school graduates with no additional degree, the unemployment rate was 5.2%; and for those without a high school diploma, the unemployment rate was 5.7%. This pattern arises because additional education typically offers better connections to the labor market and higher demand. With less attractive labor market opportunities for low-skilled workers compared to the opportunities for the more highly-skilled, including lower pay, low-skilled workers may be less motivated to find jobs.

Breaking Down Unemployment in Other Ways

The Bureau of Labor Statistics also gives information about the reasons for unemployment, as well as the length of time individuals have been unemployed. [Table 21.2](#), for example, shows the four reasons for unemployment and the percentages of the currently unemployed that fall into each category. [Table 21.3](#) shows the length of unemployment. For both of these, the data is from November 2021.(bls.gov)

Reason	Percentage
New Entrants	6.5%
Re-entrants	31.8%
Job Leavers	12.5%
Job Losers: Temporary	11.8%
Job Losers: Non Temporary	37.3%

TABLE 21.2 Reasons for Unemployment, November 2021

Length of Time	Percentage
Under 5 weeks	22.3%
5 to 14 weeks	22.3%
15 to 26 weeks	17.6%
Over 27 weeks	37.7%

TABLE 21.3 Length of Unemployment, November 2021

LINK IT UP

Watch this [speech \(http://openstax.org/l/droids\)](http://openstax.org/l/droids) on the impact of droids on the labor market.

International Unemployment Comparisons

From an international perspective, the U.S. unemployment rate typically has looked a little better than average. [Table 21.4](#) compares unemployment rates for 1991, 1996, 2001, 2006 (just before the Great Recession), and 2019 (just before the pandemic-induced recession) from several other high-income countries.

Country	1991	1996	2001	2006	2019
United States	6.8%	5.4%	4.8%	4.4%	3.7%
Canada	9.8%	8.8%	6.4%	6.2%	5.7%
Japan	2.1%	3.4%	5.1%	4.5%	2.4%
France	9.5%	12.5%	8.7%	10.1%	8.5%
Germany	5.6%	9.0%	8.9%	9.8%	3.1%
Italy	6.9%	11.7%	9.6%	7.8%	10.0%
Sweden	3.1%	9.9%	5.0%	5.2%	7.0%
United Kingdom	8.8%	8.1%	5.1%	5.5%	3.9%

TABLE 21.4 International Comparisons of Unemployment Rates

However, we need to treat cross-country comparisons of unemployment rates with care, because each country has slightly different definitions of unemployment, survey tools for measuring unemployment, and also different labor markets. For example, Japan's unemployment rates appear quite low, but Japan's economy has been mired in slow growth and recession since the late 1980s, and Japan's unemployment rate probably paints too rosy a picture of its labor market. In Japan, workers who lose their jobs are often quick to exit the labor force and not look for a new job, in which case they are not counted as unemployed. In addition, Japanese firms are often quite reluctant to fire workers, and so firms have substantial numbers of workers who are on reduced hours or officially employed, but doing very little. We can view this Japanese pattern as an unusual

method for society to provide support for the unemployed, rather than a sign of a healthy economy.

LINK IT UP

We hear about the Chinese economy in the news all the time. The value of the Chinese yuan in comparison to the U.S. dollar is likely to be part of the nightly business report, so why is the Chinese economy not included in this discussion of international unemployment? The lack of reliable statistics is the reason. This [article](http://openstax.org/l/ChinaEmployment) (<http://openstax.org/l/ChinaEmployment>) explains why.

Comparing unemployment rates in the United States and other high-income economies with unemployment rates in Latin America, Africa, Eastern Europe, and Asia is very difficult. One reason is that the statistical agencies in many poorer countries lack the resources and technical capabilities of the U.S. Bureau of the Census. However, a more difficult problem with international comparisons is that in many low-income countries, most workers are not involved in the labor market through an employer who pays them regularly. Instead, workers in these countries are engaged in short-term work, subsistence activities, and barter. Moreover, the effect of unemployment is very different in high-income and low-income countries. Unemployed workers in the developed economies have access to various government programs like unemployment insurance, welfare, and food stamps. Such programs may barely exist in poorer countries. Although unemployment is a serious problem in many low-income countries, it manifests itself in a different way than in high-income countries.

21.3 What Causes Changes in Unemployment over the Short Run

LEARNING OBJECTIVES

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

- Analyze cyclical unemployment
- Explain the relationship between sticky wages and employment using various economic arguments
- Apply supply and demand models to unemployment and wages

We have seen that unemployment varies across times and places. What causes changes in unemployment? There are different answers in the short run and in the long run. Let's look at the short run first.

Cyclical Unemployment

Let's make the plausible assumption that in the short run, from a few months to a few years, the quantity of hours that the average person is willing to work for a given wage does not change much, so the labor supply curve does not shift much. In addition, make the standard *ceteris paribus* assumption that there is no substantial short-term change in the age structure of the labor force, institutions and laws affecting the labor market, or other possibly relevant factors.

One primary determinant of the demand for labor from firms is how they perceive the state of the macro economy. If firms believe that business is expanding, then at any given wage they will desire to hire a greater quantity of labor, and the labor demand curve shifts to the right. Conversely, if firms perceive that the economy is slowing down or entering a recession, then they will wish to hire a lower quantity of labor at any given wage, and the labor demand curve will shift to the left. Economists call the variation in unemployment that the economy causes moving from expansion to recession or from recession to expansion (i.e. the business cycle) **cyclical unemployment**.

From the standpoint of the supply-and-demand model of competitive and flexible labor markets, unemployment represents something of a puzzle. In a supply-and-demand model of a labor market, as [Figure 21.5](#) illustrates, the labor market should move toward an equilibrium wage and quantity. At the equilibrium wage (W_e), the equilibrium quantity (Q_e) of labor supplied by workers should be equal to the quantity of labor demanded by employers.

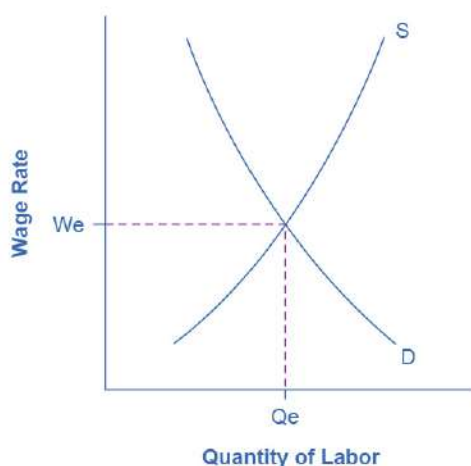


FIGURE 21.5 The Unemployment and Equilibrium in the Labor Market In a labor market with flexible wages, the equilibrium will occur at wage W_e and quantity Q_e , where the number of people who want jobs (shown by S) equals the number of jobs available (shown by D).

One possibility for unemployment is that people who are unemployed are those who are not willing to work at the current equilibrium wage, say \$10 an hour, but would be willing to work at a higher wage, like \$20 per hour. The monthly Current Population Survey would count these people as unemployed, because they say they are ready and looking for work (at \$20 per hour). However, from an economist's perspective, these people are choosing to be unemployed.

Probably a few people are unemployed because of unrealistic expectations about wages, but they do not represent the majority of the unemployed. Instead, unemployed people often have friends or acquaintances of similar skill levels who are employed, and the unemployed would be willing to work at the jobs and wages similar to what those people are receiving. However, the employers of their friends and acquaintances do not seem to be hiring. In other words, these people are involuntarily unemployed. What causes involuntary unemployment?

Why Wages Might Be Sticky Downward

If a labor market model with flexible wages does not describe unemployment very well—because it predicts that anyone willing to work at the going wage can always find a job—then it may prove useful to consider economic models in which wages are not flexible or adjust only very slowly. In particular, even though wage increases may occur with relative ease, wage decreases are few and far between.

One set of reasons why wages may be “sticky downward,” as economists put it, involves economic laws and institutions. For low-skilled workers receiving minimum wage, it is illegal to reduce their wages. For union workers operating under a multiyear contract with a company, wage cuts might violate the contract and create a labor dispute or a strike. However, minimum wages and union contracts are not a sufficient reason why wages would be sticky downward for the U.S. economy as a whole. After all, out of the 73.3 million or so employed workers in the U.S. economy who earn wages by the hour, only about 1.1 million—less than 2% of the total—do not receive compensation above the minimum wage. Similarly, labor unions represent only about 12% of American wage and salary workers. In other high-income countries, more workers may have their wages determined by unions or the minimum wage may be set at a level that applies to a larger share of workers. However, for the United States, these two factors combined affect only about 15% or less of the labor force.

Economists looking for reasons why wages might be sticky downwards have focused on factors that may characterize most labor relationships in the economy, not just a few. Many have proposed a number of different theories, but they share a common tone.

One argument is that even employees who are not union members often work under an **implicit contract**, which is that the employer will try to keep wages from falling when the economy is weak or the business is having trouble, and the employee will not expect huge salary increases when the economy or the business is strong. This wage-setting behavior acts like a form of insurance: the employee has some protection against wage declines in bad times, but pays for that protection with lower wages in good times. Clearly, this sort of implicit contract means that firms will be hesitant to cut wages, lest workers feel betrayed and work less hard or even leave the firm.

Efficiency wage theory argues that workers' productivity depends on their pay, and so employers will often find it worthwhile to pay their employees somewhat more than market conditions might dictate. One reason is that employees who receive better pay than others will be more productive because they recognize that if they were to lose their current jobs, they would suffer a decline in salary. As a result, they are motivated to work harder and to stay with the current employer. In addition, employers know that it is costly and time-consuming to hire and train new employees, so they would prefer to pay workers a little extra now rather than to lose them and have to hire and train new workers. Thus, by avoiding wage cuts, the employer minimizes costs of training and hiring new workers, and reaps the benefits of well-motivated employees.

The **adverse selection of wage cuts argument** points out that if an employer reacts to poor business conditions by reducing wages for all workers, then the best workers, those with the best employment alternatives at other firms, are the most likely to leave. The least attractive workers, with fewer employment alternatives, are more likely to stay. Consequently, firms are more likely to choose which workers should depart, through layoffs and firings, rather than trimming wages across the board. Sometimes companies that are experiencing difficult times can persuade workers to take a pay cut for the short term, and still retain most of the firm's workers. However, it is far more typical for companies to lay off some workers, rather than to cut wages for everyone.

The **insider-outsider model** of the labor force, in simple terms, argues that those already working for firms are “insiders,” while new employees, at least for a time, are “outsiders.” A firm depends on its insiders to keep the organization running smoothly, to be familiar with routine procedures, and to train new employees. However, cutting wages will alienate the insiders and damage the firm's productivity and prospects.

Finally, the **relative wage coordination argument** points out that even if most workers were hypothetically willing to see a decline in their own wages in bad economic times as long as everyone else also experiences such a decline, there is no obvious way for a decentralized economy to implement such a plan. Instead, workers confronted with the possibility of a wage cut will worry that other workers will not have such a wage cut, and so a wage cut means being worse off both in absolute terms and relative to others. As a result, workers fight hard against wage cuts.

These theories of why wages tend not to move downward differ in their logic and their implications, and figuring out the strengths and weaknesses of each theory is an ongoing subject of research and controversy among economists. All tend to imply that wages will decline only very slowly, if at all, even when the economy or a business is having tough times. When wages are inflexible and unlikely to fall, then either short-run or long-run unemployment can result. [Figure 21.6](#) illustrates this.

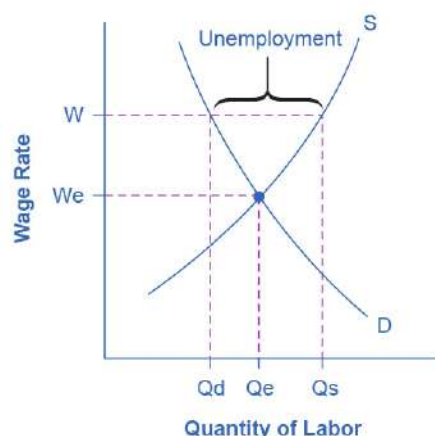


FIGURE 21.6 Sticky Wages in the Labor Market Because the wage rate is stuck at W , above the equilibrium, the number of those who want jobs (Q_s) is greater than the number of job openings (Q_d). The result is unemployment, shown by the bracket in the figure.

Figure 21.7 shows the interaction between shifts in labor demand and wages that are sticky downward. Figure 21.7 (a) illustrates the situation in which the demand for labor shifts to the right from D_0 to D_1 . In this case, the equilibrium wage rises from W_0 to W_1 and the equilibrium quantity of labor hired increases from Q_0 to Q_1 . It does not hurt employee morale at all for wages to rise.

Figure 21.7 (b) shows the situation in which the demand for labor shifts to the left, from D_0 to D_1 , as it would tend to do in a recession. Because wages are sticky downward, they do not adjust toward what would have been the new equilibrium wage (W_1), at least not in the short run. Instead, after the shift in the labor demand curve, the same quantity of workers is willing to work at that wage as before; however, the quantity of workers demanded at that wage has declined from the original equilibrium (Q_0) to Q_2 . The gap between the original equilibrium quantity (Q_0) and the new quantity demanded of labor (Q_2) represents workers who would be willing to work at the going wage but cannot find jobs. The gap represents the economic meaning of unemployment.

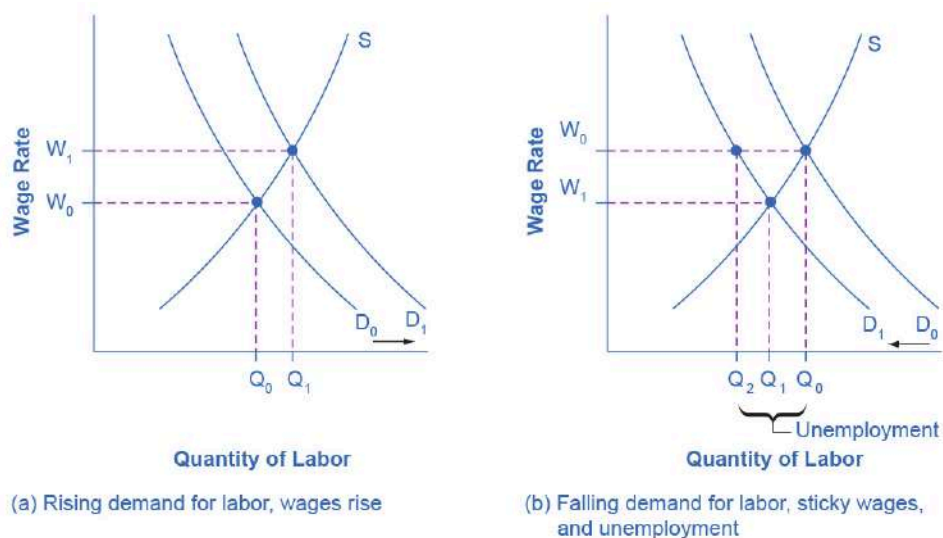


FIGURE 21.7 Rising Wage and Low Unemployment: Where Is the Unemployment in Supply and Demand? (a) In a labor market where wages are able to rise, an increase in the demand for labor from D_0 to D_1 leads to an increase in equilibrium quantity of labor hired from Q_0 to Q_1 and a rise in the equilibrium wage from W_0 to W_1 . (b) In a labor market where wages do not decline, a fall in the demand for labor from D_0 to D_1 leads to a decline in the quantity of

labor demanded at the original wage (W_0) from Q_0 to Q_2 . These workers will want to work at the prevailing wage (W_0), but will not be able to find jobs.

This analysis helps to explain the connection that we noted earlier: that unemployment tends to rise in recessions and to decline during expansions. The overall state of the economy shifts the labor demand curve and, combined with wages that are sticky downwards, unemployment changes. The rise in unemployment that occurs because of a recession is cyclical unemployment.

LINK IT UP

The St. Louis Federal Reserve Bank is the best resource for macroeconomic time series data, known as the Federal Reserve Economic Data (FRED). [FRED \(http://openstax.org/l/FRED_employment\)](http://openstax.org/l/FRED_employment) provides complete data sets on various measures of the unemployment rate as well as the monthly Bureau of Labor Statistics report on the results of the household and employment surveys.

21.4 What Causes Changes in Unemployment over the Long Run

LEARNING OBJECTIVES

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

- Explain frictional and structural unemployment
- Assess relationships between the natural rate of employment and potential real GDP, productivity, and public policy
- Identify recent patterns in the natural rate of employment
- Propose ways to combat unemployment

Cyclical unemployment explains why unemployment rises during a recession and falls during an economic expansion, but what explains the remaining level of unemployment even in good economic times? Why is the unemployment rate never zero? Even when the U.S. economy is growing strongly, the unemployment rate only rarely dips as low as 4%. Moreover, the discussion earlier in this chapter pointed out that unemployment rates in many European countries like Italy, France, and Germany have often been remarkably high at various times in the last few decades. Why does some level of unemployment persist even when economies are growing strongly? Why are unemployment rates continually higher in certain economies, through good economic years and bad? Economists have a term to describe the remaining level of unemployment that occurs even when the economy is healthy: they call it the **natural rate of unemployment**.

The Long Run: The Natural Rate of Unemployment

The natural rate of unemployment is not “natural” in the sense that water freezes at 32 degrees Fahrenheit or boils at 212 degrees Fahrenheit. It is not a physical and unchanging law of nature. Instead, it is only the “natural” rate because it is the unemployment rate that would result from the combination of economic, social, and political factors that exist at a time—assuming the economy was neither booming nor in recession. These forces include the usual pattern of companies expanding and contracting their workforces in a dynamic economy, social and economic forces that affect the labor market, or public policies that affect either the eagerness of people to work or the willingness of businesses to hire. Let’s discuss these factors in more detail.

Frictional Unemployment

In a market economy, some companies are always going broke for a variety of reasons: old technology; poor management; good management that happened to make bad decisions; shifts in tastes of consumers so that less of the firm’s product is desired; a large customer who went broke; or tough domestic or foreign competitors. Conversely, other companies will be doing very well for just the opposite reasons and looking to hire more employees. In a perfect world, all of those who lost jobs would immediately find new ones. However, in the real world, even if the number of job seekers is equal to the number of job vacancies, it takes time to find out about new jobs, to interview and figure out if the new job is a good match, or perhaps to sell a house and

buy another in proximity to a new job. Economists call the unemployment that occurs in the meantime, as workers move between jobs, **frictional unemployment**. Frictional unemployment is not inherently a bad thing. It takes time on part of both the employer and the individual to match those looking for employment with the correct job openings. For individuals and companies to be successful and productive, you want people to find the job for which they are best suited, not just the first job offered.

In the mid-2000s, before the 2008–2009 recession, it was true that about 7% of U.S. workers saw their jobs disappear in any three-month period. However, in periods of economic growth, these destroyed jobs are counterbalanced for the economy as a whole by a larger number of jobs created. In 2019, for example, there were typically about 6 million unemployed people at any given time in the U.S. economy. Even though about two-thirds of those unemployed people found a job in 14 weeks or fewer, the unemployment rate did not change much during the year, because those who found new jobs were largely offset by others who lost jobs.

Of course, it would be preferable if people who were losing jobs could immediately and easily move into newly created jobs, but in the real world, that is not possible. Someone who is laid off by a textile mill in South Carolina cannot turn around and immediately start working for a textile mill in California. Instead, the adjustment process happens in ripples. Some people find new jobs near their old ones, while others find that they must move to new locations. Some people can do a very similar job with a different company, while others must start new career paths. Some people may be near retirement and decide to look only for part-time work, while others want an employer that offers a long-term career path. The frictional unemployment that results from people moving between jobs in a dynamic economy may account for one to two percentage points of total unemployment.

The level of frictional unemployment will depend on how easy it is for workers to learn about alternative jobs, which may reflect the ease of communications about job prospects in the economy. The extent of frictional unemployment will also depend to some extent on how willing people are to move to new areas to find jobs—which in turn may depend on history and culture.

Frictional unemployment and the natural rate of unemployment also seem to depend on the age distribution of the population. [Figure 21.4](#) (b) showed that unemployment rates are typically lower for people between 25–54 years of age or aged 55 and over than they are for those who are younger. “Prime-age workers,” as those in the 25–54 age bracket are sometimes called, are typically at a place in their lives when they want to have a job and income arriving at all times. In addition, older workers who lose jobs may prefer to opt for retirement. By contrast, it is likely that a relatively high proportion of those who are under 25 will be trying out jobs and life options, and this leads to greater job mobility and hence higher frictional unemployment. Thus, a society with a relatively high proportion of young workers, like the U.S. beginning in the mid-1960s when Baby Boomers began entering the labor market, will tend to have a higher unemployment rate than a society with a higher proportion of its workers in older ages.

Structural Unemployment

Another factor that influences the natural rate of unemployment is the amount of **structural unemployment**. The structurally unemployed are individuals who have no jobs because they lack skills valued by the labor market, either because demand has shifted away from the skills they do have, or because they never learned any skills. An example of the former would be the unemployment among aerospace engineers after the U.S. space program downsized in the 1970s. An example of the latter would be high school dropouts.

Some people worry that technology causes structural unemployment. In the past, new technologies have put lower skilled employees out of work, but at the same time they create demand for higher skilled workers to use the new technologies. Education seems to be the key in minimizing the amount of structural unemployment. Individuals who have degrees can be retrained if they become structurally unemployed. For people with no skills and little education, that option is more limited.

Natural Unemployment and Potential Real GDP

The natural unemployment rate is related to two other important concepts: full employment and potential real GDP. Economists consider the economy to be at full employment when the actual unemployment rate is equal to the natural unemployment rate. When the economy is at full employment, real GDP is equal to potential real GDP. By contrast, when the economy is below full employment, the unemployment rate is greater than the natural unemployment rate and real GDP is less than potential. Finally, when the economy is above full employment, then the unemployment rate is less than the natural unemployment rate and real GDP is greater than potential. Operating above potential is only possible for a short while, since it is analogous to all workers working overtime.

Productivity Shifts and the Natural Rate of Unemployment

Unexpected shifts in productivity can have a powerful effect on the natural rate of unemployment. Over time, workers' productivity determines the level of wages in an economy. After all, if a business paid workers more than could be justified by their productivity, the business will ultimately lose money and go bankrupt. Conversely, if a business tries to pay workers less than their productivity then, in a competitive labor market, other businesses will find it worthwhile to hire away those workers and pay them more.

However, adjustments of wages to productivity levels will not happen quickly or smoothly. Employers typically review wages only once or twice a year. In many modern jobs, it is difficult to measure productivity at the individual level. For example, how precisely would one measure the quantity produced by an accountant who is one of many people working in the tax department of a large corporation? Because productivity is difficult to observe, employers often determine wage increases based on recent experience with productivity. If productivity has been rising at, say, 2% per year, then wages rise at that level as well. However, when productivity changes unexpectedly, it can affect the natural rate of unemployment for a time.

The U.S. economy in the 1970s and 1990s provides two vivid examples of this process. In the 1970s, productivity growth slowed down unexpectedly (as we discussed in [Economic Growth](#)). For example, output per hour of U.S. workers in the business sector increased at an annual rate of 3.3% per year from 1960 to 1973, but only 0.8% from 1973 to 1982. [Figure 21.8](#) (a) illustrates the situation where the demand for labor—that is, the quantity of labor that business is willing to hire at any given wage—has been shifting out a little each year because of rising productivity, from D_0 to D_1 to D_2 . As a result, equilibrium wages have been rising each year from W_0 to W_1 to W_2 . However, when productivity unexpectedly slows down, the pattern of wage increases does not adjust right away. Wages keep rising each year from W_2 to W_3 to W_4 , but the demand for labor is no longer shifting up. A gap opens where the quantity of labor supplied at wage level W_4 is greater than the quantity demanded. The natural rate of unemployment rises. In the aftermath of this unexpectedly low productivity in the 1970s, the national unemployment rate did not fall below 7% from May, 1980 until 1986. Over time, the rise in wages will adjust to match the slower gains in productivity, and the unemployment rate will ease back down, but this process may take years.

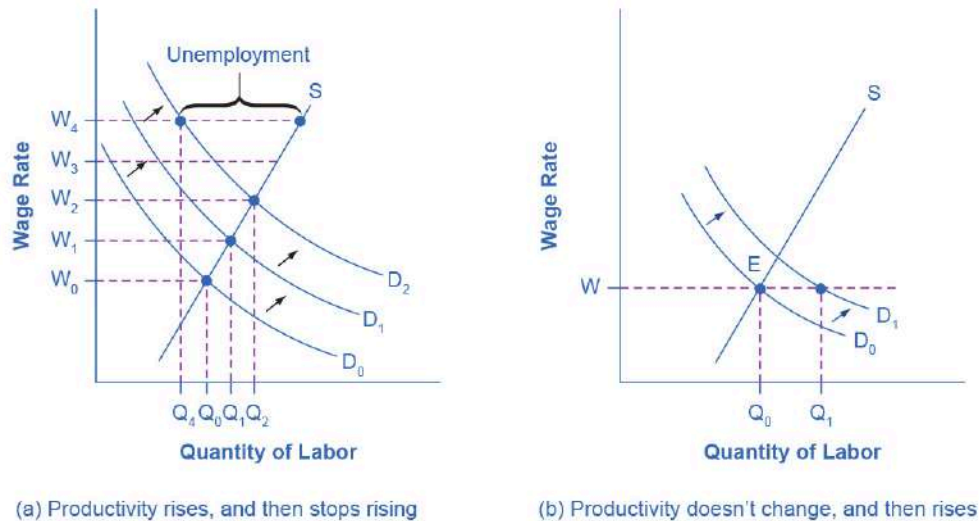


FIGURE 21.8 Unexpected Productivity Changes and Unemployment (a) Productivity is rising, increasing the demand for labor. Employers and workers become used to the pattern of wage increases. Then productivity suddenly stops increasing. However, the expectations of employers and workers for wage increases do not shift immediately, so wages keep rising as before. However, the demand for labor has not increased, so at wage W_4 , unemployment exists where the quantity supplied of labor exceeds the quantity demanded. (b) The rate of productivity increase has been zero for a time, so employers and workers have come to accept the equilibrium wage level (W). Then productivity increases unexpectedly, shifting demand for labor from D_0 to D_1 . At the wage (W), this means that the quantity demanded of labor exceeds the quantity supplied, and with job offers plentiful, the unemployment rate will be low.

The late 1990s provide an opposite example: instead of the surprise decline in productivity that occurred in the 1970s, productivity unexpectedly rose in the mid-1990s. The annual growth rate of real output per hour of labor increased from 1.7% from 1980–1995, to an annual rate of 2.6% from 1995–2001. Let's simplify the situation a bit, so that the economic lesson of the story is easier to see graphically, and say that productivity had not been increasing at all in earlier years, so the intersection of the labor market was at point E in Figure 21.8 (b), where the demand curve for labor (D_0) intersects the supply curve for labor. As a result, real wages were not increasing. Now, productivity jumps upward, which shifts the demand for labor out to the right, from D_0 to D_1 . At least for a time, however, wages are still set according to the earlier expectations of no productivity growth, so wages do not rise. The result is that at the prevailing wage level (W), the quantity of labor demanded (Q_d) will for a time exceed the quantity of labor supplied (Q_s), and unemployment will be very low—actually below the natural level of unemployment for a time. This pattern of unexpectedly high productivity helps to explain why the unemployment rate stayed below 4.5%—quite a low level by historical standards—from 1998 until after the U.S. economy had entered a recession in 2001.

Levels of unemployment will tend to be somewhat higher on average when productivity is unexpectedly low, and conversely, will tend to be somewhat lower on average when productivity is unexpectedly high. However, over time, wages do eventually adjust to reflect productivity levels.

Public Policy and the Natural Rate of Unemployment

Public policy can also have a powerful effect on the natural rate of unemployment. On the supply side of the labor market, public policies to assist the unemployed can affect how eager people are to find work. For example, if a worker who loses a job is guaranteed a generous package of unemployment insurance, welfare benefits, food stamps, and government medical benefits, then the opportunity cost of unemployment is lower and that worker will be less eager to seek a new job.

What seems to matter most is not just the amount of these benefits, but how long they last. A society that

provides generous help for the unemployed that cuts off after, say, six months, may provide less of an incentive for unemployment than a society that provides less generous help that lasts for several years. Conversely, government assistance for job search or retraining can in some cases encourage people back to work sooner. See the Clear it Up to learn how the U.S. handles unemployment insurance.



CLEAR IT UP

How does U.S. unemployment insurance work?

Unemployment insurance is a joint federal–state program that the federal government enacted in 1935. While the federal government sets minimum standards for the program, state governments conduct most of the administration.

The funding for the program is a federal tax collected from employers. The federal government requires tax collection on the first \$7,000 in wages paid to each worker; however, states can choose to collect the tax on a higher amount if they wish, and 41 states have set a higher limit. States can choose the length of time that they pay benefits, although most states limit unemployment benefits to 26 weeks—with extensions possible in times of especially high unemployment. The states then use the fund to pay benefits to those who become unemployed. Average unemployment benefits are equal to about one-third of the wage that the person earned in their previous job, but the level of unemployment benefits varies considerably across states.

Bottom 10 States That Pay the Lowest Benefit per Week		Top 10 States That Pay the Highest Benefit per Week	
Michigan	\$362	Washington	\$929
North Carolina	\$350	Massachusetts	\$823
South Carolina	\$326	Minnesota	\$740
Missouri	\$320	New Jersey	\$713
Florida	\$275	Connecticut	\$649
Tennessee	\$275	Oregon	\$648
Alabama	\$275	Hawaii	\$648
Louisiana	\$275	North Dakota	\$618
Arizona	\$240	Colorado	\$618
Mississippi	\$235	Rhode Island	\$586

TABLE 21.5 Maximum Weekly Unemployment Benefits by State in 2021 (Source: <http://www.savingtoinvest.com/maximum-weekly-unemployment-benefits-by-state/>)

One other interesting thing to note about the classifications of unemployment—an individual does not have to collect unemployment benefits to be classified as unemployed. While there are statistics kept and studied relating to how many people are collecting unemployment insurance, this is not the source of unemployment rate information.



LINK IT UP

View this [article \(http://openstax.org/l/NYT_Benefits\)](http://openstax.org/l/NYT_Benefits) for an explanation of exactly who is eligible for

unemployment benefits.

On the demand side of the labor market, government rules, social institutions, and the presence of unions can affect the willingness of firms to hire. For example, if a government makes it hard for businesses to start up or to expand, by wrapping new businesses in bureaucratic red tape, then businesses will become more discouraged about hiring. Government regulations can make it harder to start a business by requiring that a new business obtain many permits and pay many fees, or by restricting the types and quality of products that a company can sell. Other government regulations, like zoning laws, may limit where companies can conduct business, or whether businesses are allowed to be open during evenings or on Sunday.

Whatever defenses may be offered for such laws in terms of social value—like the value some Christians place on not working on Sunday, or Orthodox Jews or highly observant Muslims on Saturday—these kinds of restrictions impose a barrier between some willing workers and other willing employers, and thus contribute to a higher natural rate of unemployment. Similarly, if government makes it difficult to fire or lay off workers, businesses may react by trying not to hire more workers than strictly necessary—since laying these workers off would be costly and difficult. High minimum wages may discourage businesses from hiring low-skill workers. Government rules may encourage and support powerful unions, which can then push up wages for union workers, but at a cost of discouraging businesses from hiring those workers.

The Natural Rate of Unemployment in Recent Years

The underlying economic, social, and political factors that determine the natural rate of unemployment can change over time, which means that the natural rate of unemployment can change over time, too.

Estimates by economists of the natural rate of unemployment in the U.S. economy in the early 2000s run at about 4.5 to 5.5%. This is a lower estimate than earlier. We outline three of the common reasons that economists propose for this change below.

1. The internet has provided a remarkable new tool through which job seekers can find out about jobs at different companies and can make contact with relative ease. An internet search is far easier than trying to find a list of local employers and then hunting up phone numbers for all of their human resources departments, and requesting a list of jobs and application forms. Social networking sites such as LinkedIn have changed how people find work as well.
2. The growth of the temporary worker industry has probably helped to reduce the natural rate of unemployment. In the early 1980s, only about 0.5% of all workers held jobs through temp agencies. By the early 2000s, the figure had risen above 2%. Temp agencies can provide jobs for workers while they are looking for permanent work. They can also serve as a clearinghouse, helping workers find out about jobs with certain employers and getting a tryout with the employer. For many workers, a temp job is a stepping-stone to a permanent job that they might not have heard about or obtained any other way, so the growth of temp jobs will also tend to reduce frictional unemployment.
3. The aging of the “baby boom generation”—the especially large generation of Americans born between 1946 and 1964—meant that the proportion of young workers in the economy was relatively high in the 1970s, as the boomers entered the labor market, but is relatively low today. As we noted earlier, middle-aged and older workers are far more likely to experience low unemployment than younger workers, a factor that tends to reduce the natural rate of unemployment as the baby boomers age.

The combined result of these factors is that the natural rate of unemployment was on average lower in the 1990s and the early 2000s than in the 1980s. The 2008–2009 Great Recession pushed monthly unemployment rates up to 10% in late 2009. However, even at that time, the Congressional Budget Office was forecasting that by 2015, unemployment rates would fall back to about 5%. During the first two months of 2020, the unemployment rate held steady at 3.5%. As of the first quarter of 2022, the Congressional Budget Office estimates the natural rate to be 4.6%.

The Natural Rate of Unemployment in Europe

By the standards of other high-income economies, the natural rate of unemployment in the U.S. economy appears relatively low. Through good economic years and bad, many European economies have had unemployment rates hovering near 10%, or even higher, since the 1970s. European rates of unemployment have been higher not because recessions in Europe have been deeper, but rather because the conditions underlying supply and demand for labor have been different in Europe, in a way that has created a much higher natural rate of unemployment.

Many European countries have a combination of generous welfare and unemployment benefits, together with nests of rules that impose additional costs on businesses when they hire. In addition, many countries have laws that require firms to give workers months of notice before laying them off and to provide substantial severance or retraining packages after laying them off. The legally required notice before laying off a worker can be more than three months in Spain, Germany, Denmark, and Belgium, and the legally required severance package can be as high as a year's salary or more in Austria, Spain, Portugal, Italy, and Greece. Such laws will surely discourage laying off or firing current workers. However, when companies know that it will be difficult to fire or lay off workers, they also become hesitant about hiring in the first place.

We can attribute the typically higher levels of unemployment in many European countries in recent years, which have prevailed even when economies are growing at a solid pace, to the fact that the sorts of laws and regulations that lead to a high natural rate of unemployment are much more prevalent in Europe than in the United States.

A Preview of Policies to Fight Unemployment

The [Government Budgets and Fiscal Policy](#) and [Macroeconomic Policy Around the World](#) chapters provide a detailed discussion of how to fight unemployment, when we can discuss these policies in the context of the full array of macroeconomic goals and frameworks for analysis. However, even at this preliminary stage, it is useful to preview the main issues concerning policies to fight unemployment.

The remedy for unemployment will depend on the diagnosis. Cyclical unemployment is a short-term problem, caused because the economy is in a recession. Thus, the preferred solution will be to avoid or minimize recessions. As [Government Budgets and Fiscal Policy](#) discusses, governments can enact this policy by stimulating the overall buying power in the economy, so that firms perceive that sales and profits are possible, which makes them eager to hire.

Dealing with the natural rate of unemployment is trickier. In a market-oriented economy, firms will hire and fire workers. Governments cannot control this. Furthermore, the evolving age structure of the economy's population, or unexpected shifts in productivity are beyond a government's control and, will affect the natural rate of unemployment for a time. However, as the example of high ongoing unemployment rates for many European countries illustrates, government policy clearly can affect the natural rate of unemployment that will persist even when GDP is growing.

When a government enacts policies that will affect workers or employers, it must examine how these policies will affect the information and incentives employees and employers have to find one another. For example, the government may have a role to play in helping some of the unemployed with job searches. Governments may need to rethink the design of their programs that offer assistance to unemployed workers and protections to employed workers so that they will not unduly discourage the supply of labor. Similarly, governments may need to reassess rules that make it difficult for businesses to begin or to expand so that they will not unduly discourage the demand for labor. The message is not that governments should repeal all laws affecting labor markets, but only that when they enact such laws, a society that cares about unemployment will need to consider the tradeoffs involved.



BRING IT HOME

Unemployment and the COVID-19 Pandemic

Almost two years after the pandemic began, the unemployment rate was on track to go below 4% again by early- to mid-2022. While this is great news for workers, we have also seen that the story of the labor market is more complicated than a single statistic might suggest. Millions remained out of the labor force due to the public health situation, and the percentage of workers unemployed for longer than 26 weeks was still high.

The shift to remote work helped the unemployment rate come down by providing greater flexibility for workers concerned with their health and safety. But it didn't help workers, especially women, who continued to be burdened with excessive care responsibilities at home. During 2020, the unemployment rate for women exceeded that of men by over a full percentage point. Additionally, virus variants in other countries threatened to disrupt progress made at home and abroad.

The pandemic has helped shed light on how we can use information from all areas of the labor market to evaluate the health of the economy.

Key Terms

- adverse selection of wage cuts argument** if employers reduce wages for all workers, the best will leave
- cyclical unemployment** unemployment closely tied to the business cycle, like higher unemployment during a recession
- discouraged workers** those who have stopped looking for employment due to the lack of suitable positions available
- efficiency wage theory** the theory that the productivity of workers, either individually or as a group, will increase if the employer pays them more
- frictional unemployment** unemployment that occurs as workers move between jobs
- implicit contract** an unwritten agreement in the labor market that the employer will try to keep wages from falling when the economy is weak or the business is having trouble, and the employee will not expect huge salary increases when the economy or the business is strong
- insider-outsider model** those already working for the firm are “insiders” who know the procedures; the other workers are “outsiders” who are recent or prospective hires
- labor force participation rate** this is the percentage of adults in an economy who are either employed or who are unemployed and looking for a job
- natural rate of unemployment** the unemployment rate that would exist in a growing and healthy economy from the combination of economic, social, and political factors that exist at a given time
- out of the labor force** those who are not working and not looking for work—whether they want employment or not; also termed “not in the labor force”
- relative wage coordination argument** across-the-board wage cuts are hard for an economy to implement, and workers fight against them
- structural unemployment** unemployment that occurs because individuals lack skills valued by employers
- underemployed** individuals who are employed in a job that is below their skills
- unemployment rate** the percentage of adults who are in the labor force and thus seeking jobs, but who do not have jobs

Key Concepts and Summary

21.1 How Economists Define and Compute Unemployment Rate

Unemployment imposes high costs. Unemployed individuals experience loss of income and stress. An economy with high unemployment suffers an opportunity cost of unused resources. We can divide the adult population into those in the labor force and those out of the labor force. In turn, we divide those in the labor force into employed and unemployed. A person without a job must be willing and able to work and actively looking for work to be counted as unemployed; otherwise, a person without a job is counted as out of the labor force. Economists define the unemployment rate as the number of unemployed persons divided by the number of persons in the labor force (not the overall adult population). The Current Population Survey (CPS) conducted by the United States Census Bureau measures the percentage of the labor force that is unemployed. The establishment payroll survey by the Bureau of Labor Statistics measures the net change in jobs created for the month.

21.2 Patterns of Unemployment

The U.S. unemployment rate rises during periods of recession and depression, but falls back to the range of 4% to 6% when the economy is strong. The unemployment rate never falls to zero. Despite enormous growth in the size of the U.S. population and labor force in the twentieth century, along with other major trends like globalization and new technology, the unemployment rate shows no long-term rising trend.

Unemployment rates differ by group: higher for African-Americans and Hispanic people than for White people; higher for less educated than more educated; higher for the young than the middle-aged. Women's unemployment rates used to be higher than men's, but in recent years men's and women's unemployment

rates have been very similar. In recent years, unemployment rates in the United States have compared favorably with unemployment rates in most other high-income economies.

21.3 What Causes Changes in Unemployment over the Short Run

Cyclical unemployment rises and falls with the business cycle. In a labor market with flexible wages, wages will adjust in such a market so that quantity demanded of labor always equals the quantity supplied of labor at the equilibrium wage. Economists have proposed many theories for why wages might not be flexible, but instead may adjust only in a “sticky” way, especially when it comes to downward adjustments: implicit contracts, efficiency wage theory, adverse selection of wage cuts, insider-outsider model, and relative wage coordination.

21.4 What Causes Changes in Unemployment over the Long Run

The natural rate of unemployment is the rate of unemployment that the economic, social, and political forces in the economy would cause even when the economy is not in a recession. These factors include the frictional unemployment that occurs when people either choose to change jobs or are put out of work for a time by the shifts of a dynamic and changing economy. They also include any laws concerning conditions of hiring and firing that have the undesired side effect of discouraging job formation. They also include structural unemployment, which occurs when demand shifts permanently away from a certain type of job skill.

Self-Check Questions

1. Suppose the adult population over the age of 16 is 237.8 million and the labor force is 153.9 million (of whom 139.1 million are employed). How many people are “not in the labor force?” What are the proportions of employed, unemployed and not in the labor force in the population? *Hint:* Proportions are percentages.
2. Using the above data, what is the unemployment rate? These data are U.S. statistics from 2010. How does it compare to the February 2015 unemployment rate computed earlier?
3. Over the long term, has the U.S. unemployment rate generally trended up, trended down, or remained at basically the same level?
4. Whose unemployment rates are commonly higher in the U.S. economy:
 - a. White or non-White people?
 - b. The young or the middle-aged?
 - c. College graduates or high school graduates?
5. Beginning in the 1970s and continuing for three decades, women entered the U.S. labor force in a big way. If we assume that wages are sticky in a downward direction, but that around 1970 the demand for labor equaled the supply of labor at the current wage rate, what do you imagine happened to the wage rate, employment, and unemployment as a result of increased labor force participation?
6. Is the increase in labor force participation rates among women better thought of as causing an increase in cyclical unemployment or an increase in the natural rate of unemployment? Why?
7. Many college students graduate from college before they have found a job. When graduates begin to look for a job, they are counted as what category of unemployed?

Review Questions

8. What is the difference between being unemployed and being out of the labor force?
9. How do you calculate the unemployment rate? How do you calculate the labor force participation rate?
10. Are all adults who do not hold jobs counted as unemployed?

11. If you are out of school but working part time, are you considered employed or unemployed in U.S. labor statistics? If you are a full time student and working 12 hours a week at the college cafeteria are you considered employed or not in the labor force? If you are a senior citizen who is collecting social security and a pension and working as a greeter at Wal-Mart are you considered employed or not in the labor force?
12. What happens to the unemployment rate when unemployed workers are reclassified as discouraged workers?
13. What happens to the labor force participation rate when employed individuals are reclassified as unemployed? What happens when they are reclassified as discouraged workers?
14. What are some of the problems with using the unemployment rate as an accurate measure of overall joblessness?
15. What criteria do the BLS use to count someone as employed? As unemployed?
16. Assess whether the following would be counted as “unemployed” in the Current Employment Statistics survey.
 - a. A husband willingly stays home with children while his wife works.
 - b. A manufacturing worker whose factory just closed down.
 - c. A college student doing an unpaid summer internship.
 - d. A retiree.
 - e. Someone who has been out of work for two years but keeps looking for a job.
 - f. Someone who has been out of work for two months but isn’t looking for a job.
 - g. Someone who hates her present job and is actively looking for another one.
 - h. Someone who decides to take a part time job because she could not find a full time position.
17. Are U.S. unemployment rates typically higher, lower, or about the same as unemployment rates in other high-income countries?
18. Are U.S. unemployment rates distributed evenly across the population?
19. When would you expect cyclical unemployment to be rising? Falling?
20. Why is there unemployment in a labor market with flexible wages?
21. Name and explain some of the reasons why wages are likely to be sticky, especially in downward adjustments.
22. What term describes the remaining level of unemployment that occurs even when the economy is healthy?
23. What forces create the natural rate of unemployment for an economy?
24. Would you expect the natural rate of unemployment to be roughly the same in different countries?
25. Would you expect the natural rate of unemployment to remain the same within one country over the long run of several decades?
26. What is frictional unemployment? Give examples of frictional unemployment.
27. What is structural unemployment? Give examples of structural unemployment.
28. After several years of economic growth, would you expect the unemployment in an economy to be mainly cyclical or mainly due to the natural rate of unemployment? Why?

29. What type of unemployment (cyclical, frictional, or structural) applies to each of the following:
- landscapers laid off in response to a drop in new housing construction during a recession.
 - coal miners laid off due to EPA regulations that shut down coal fired power
 - a financial analyst who quits his/her job in Chicago and is pursuing similar work in Arizona
 - printers laid off due to drop in demand for printed catalogues and flyers as firms go the internet to promote an advertise their products.
 - factory workers in the U.S. laid off as the plants shut down and move to Mexico and Ireland.

Critical Thinking Questions

30. Using the definition of the unemployment rate, is an increase in the unemployment rate necessarily a bad thing for a nation?
31. Is a decrease in the unemployment rate necessarily a good thing for a nation? Explain.
32. If many workers become discouraged from looking for jobs, explain how the number of jobs could decline but the unemployment rate could fall at the same time.
33. Would you expect hidden unemployment to be higher, lower, or about the same when the unemployment rate is high, say 10%, versus low, say 4%? Explain.
34. Is the higher unemployment rates for minority workers necessarily an indication of discrimination? What could be some other reasons for the higher unemployment rate?
35. While unemployment is highly negatively correlated with the level of economic activity, in the real world it responds with a lag. In other words, firms do not immediately lay off workers in response to a sales decline. They wait a while before responding. Similarly, firms do not immediately hire workers when sales pick up. What do you think accounts for the lag in response time?
36. Why do you think that unemployment rates are lower for individuals with more education?
37. Do you think it is rational for workers to prefer sticky wages to wage cuts, when the consequence of sticky wages is unemployment for some workers? Why or why not? How do the reasons for sticky wages explained in this section apply to your argument?
38. Under what condition would a decrease in unemployment be bad for the economy?
39. Under what condition would an increase in the unemployment rate be a positive sign?
40. As the baby boom generation retires, the ratio of retirees to workers will increase noticeably. How will this affect the Social Security program? How will this affect the standard of living of the average American?
41. Unemployment rates have been higher in many European countries in recent decades than in the United States. Is the main reason for this long-term difference in unemployment rates more likely to be cyclical unemployment or the natural rate of unemployment? Explain briefly.
42. Is it desirable to pursue a goal of zero unemployment? Why or why not?
43. Is it desirable to eliminate natural unemployment? Why or why not? *Hint:* Think about what our economy would look like today and what assumptions would have to be met to have a zero rate of natural unemployment.
44. The U.S. unemployment rate increased from 4.6% in July 2001 to 5.9% by June 2002. Without studying the subject in any detail, would you expect that a change of this kind is more likely to be due to cyclical unemployment or a change in the natural rate of unemployment? Why?

Problems

45. A country with a population of eight million adults has five million employed, 500,000 unemployed, and the rest of the adult population is out of the labor force. What's the unemployment rate? What share of population is in the labor force? Sketch a pie chart that divides the adult population into these three groups.
46. A government passes a family-friendly law that no companies can have evening, nighttime, or weekend hours, so that everyone can be home with their families during these times. Analyze the effect of this law using a demand and supply diagram for the labor market: first assuming that wages are flexible, and then assuming that wages are sticky downward.
47. As the baby boomer generation retires, what should happen to wages and employment? Can you show this graphically?



FIGURE 22.1 Big Bucks in Zimbabwe This bill was worth 100 billion Zimbabwean dollars when issued in 2008. There were even bills issued with a face value of 100 trillion Zimbabwean dollars. The bills had \$100,000,000,000,000 written on them. Unfortunately, they were almost worthless. At one point, 621,984,228 Zimbabwean dollars were equal to one U.S. dollar. Eventually, the country abandoned its own currency and allowed people to use foreign currency for purchases. (Credit: modification of "100 Billion Dollars" by Peat Bakke/Flickr, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Tracking Inflation
- How to Measure Changes in the Cost of Living
- How the U.S. and Other Countries Experience Inflation
- The Confusion Over Inflation
- Indexing and Its Limitations

Introduction to Inflation



BRING IT HOME

A \$550 Million Loaf of Bread?

If you were born within the last three decades in the United States, Canada, or many other countries in the developed world, you probably have no real experience with a high rate of inflation. Inflation is when most prices in an entire economy are rising. However, there is an extreme form of inflation called hyperinflation. This occurred in Germany between 1921 and 1928, and more recently in Zimbabwe between 2008 and 2009. In November 2008, Zimbabwe had an inflation rate of 79.6 billion percent. In contrast, in 2014, the United States had an average annual rate of inflation of 1.6%.

Zimbabwe's inflation rate was so high it is difficult to comprehend, so let's put it into context. It is equivalent to price increases of 98% per day. This means that, from one day to the next, prices essentially double. What is life like in an economy afflicted with hyperinflation? Most of you reading this will have never experienced this phenomenon. The government adjusted prices for commodities in Zimbabwean dollars several times *each day*. There was no desire to hold on to currency since it lost value by the minute. The people there spent a great deal of time getting rid of any cash they acquired by purchasing whatever food or other commodities they could find. At one point, a loaf of bread cost 550 million Zimbabwean dollars. Teachers' salaries were in the trillions a month; however, this was equivalent to only one U.S. dollar a day. At its height, it took 621,984,228 Zimbabwean dollars to purchase one U.S. dollar.

Government agencies had no money to pay their workers so they started printing money to pay their bills rather than raising taxes. Rising prices caused the government to enact price controls on private businesses, which led to shortages and the emergence of black markets. In 2009, the country abandoned its currency and allowed people to use foreign currencies for purchases.

How does this happen? How can both government and the economy fail to function at the most basic level? Before we consider these extreme cases of hyperinflation, let's first look at inflation itself.

Inflation is a general and ongoing rise in the level of prices in an entire economy. Inflation does not refer to a change in relative prices. A relative price change occurs when you see that the price of tuition has risen, but the price of laptops has fallen. Inflation, on the other hand, means that there is pressure for prices to rise in most markets in the economy. In addition, price increases in the supply-and-demand model were one-time events, representing a shift from a previous equilibrium to a new one. Inflation implies an ongoing rise in prices. If inflation happened for one year and then stopped, then it would not be inflation any more.

This chapter begins by showing how to combine prices of individual goods and services to create a measure of overall inflation. It discusses the historical and recent experience of inflation, both in the United States and in other countries around the world. Other chapters have sometimes included a note under an exhibit or a parenthetical reminder in the text saying that the numbers have been adjusted for inflation. In this chapter, it is time to show how to use inflation statistics to adjust other economic variables, so that you can tell how much of, for example, we can attribute the rise in GDP over different periods of time to an actual increase in the production of goods and services and how much we should attribute to the fact that prices for most items have risen.

Inflation has consequences for people and firms throughout the economy, in their roles as lenders and borrowers, wage-earners, taxpayers, and consumers. The chapter concludes with a discussion of some imperfections and biases in the inflation statistics, and a preview of policies for fighting inflation that we will discuss in other chapters.

22.1 Tracking Inflation

LEARNING OBJECTIVES

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

- Calculate the annual rate of inflation
- Explain and use index numbers and base years when simplifying the total quantity spent over a year for products
- Calculate inflation rates using index numbers

Dinner table conversations where you might have heard about inflation usually entail reminiscing about when “everything seemed to cost so much less. You used to be able to buy three gallons of gasoline for a dollar and then go see an afternoon movie for another dollar.” [Table 22.1](#) compares some prices of common goods in 1970 and 2017. Of course, the average prices in this table may not reflect the prices where you live. The cost of living in New York City is much higher than in Houston, Texas, for example. In addition, certain products have evolved over recent decades. A new car in 2021, loaded with antipollution equipment, safety gear,

computerized engine controls, and many other technological advances, is a more advanced machine (and more fuel efficient) than your typical 1970s car. However, put details like these to one side for the moment, and look at the overall pattern. The primary reason behind the price rises in [Table 22.1](#)—and all the price increases for the other products in the economy—is not specific to the market for housing or cars or gasoline or movie tickets. Instead, it is part of a general rise in the level of all prices. At the beginning of 2021, \$1 had about the same purchasing power in overall terms of goods and services as 15 cents did in 1972, because of the amount of inflation that has occurred over that time period.

Items	1970	2021
Pound of ground beef	\$0.66	\$5.96
Pound of butter	\$0.87	\$3.50
Movie ticket	\$1.55	\$13.70
Sales price of new home (median)	\$22,000	\$408,800
New car	\$3,000	\$42,000
Gallon of gasoline	\$0.36	\$3.32
Average hourly wage for a manufacturing worker	\$3.23	\$30.11
Per capita GDP	\$5,069	\$63,543

TABLE 22.1 Price Comparisons, 1970 and 2021 (Sources: See chapter References at end of book.)

Moreover, the power of inflation does not affect just goods and services, but wages and income levels, too. The second-to-last row of [Table 22.1](#) shows that the average hourly wage for a manufacturing worker increased nearly ten-fold from 1970 to 2021. The average worker in 2021 is better educated and more productive than the average worker in 1970—but not six times more productive. Per capita GDP increased substantially from 1970 to 2021, but is the average person in the U.S. economy really more than twelve times better off in just 51 years? Not likely.

A modern economy has millions of goods and services whose prices are continually quivering in the breezes of supply and demand. How can all of these shifts in price attribute to a single inflation rate? As with many problems in economic measurement, the conceptual answer is reasonably straightforward: Economists combine prices of a variety of goods and services into a single price level. The inflation rate is simply the percentage change in the price level. Applying the concept, however, involves some practical difficulties.

The Price of a Basket of Goods

To calculate the price level, economists begin with the concept of a **basket of goods and services**, consisting of the different items individuals, businesses, or organizations typically buy. The next step is to look at how the prices of those items change over time. In thinking about how to combine individual prices into an overall price level, many people find that their first impulse is to calculate the average of the prices. Such a calculation, however, could easily be misleading because some products matter more than others.

Changes in the prices of goods for which people spend a larger share of their incomes will matter more than changes in the prices of goods for which people spend a smaller share of their incomes. For example, an increase of 10% in the rental rate on housing matters more to most people than whether the price of carrots rises by 10%. To construct an overall measure of the price level, economists compute a weighted average of the

prices of the items in the basket, where the weights are based on the actual quantities of goods and services people buy. The following Work It Out feature walks you through the steps of calculating the annual rate of inflation based on a few products.

WORK IT OUT

Calculating an Annual Rate of Inflation

Consider the simple basket of goods with only three items, represented in [Table 22.2](#). Say that in any given month, a college student spends money on 20 hamburgers, one bottle of aspirin, and five movies. The table provides prices for these items over four years through each time period (Pd). Prices of some goods in the basket may rise while others fall. In this example, the price of aspirin does not change over the four years, while movies increase in price and hamburgers bounce up and down. The table shows the cost of buying the given basket of goods at the prices prevailing at that time.

Items	Hamburger	Aspirin	Movies	Total	Inflation Rate
Qty	20	1 bottle	5	-	-
(Pd 1) Price	\$3.00	\$10.00	\$6.00	-	-
(Pd 1) Amount Spent	\$60.00	\$10.00	\$30.00	\$100.00	-
(Pd 2) Price	\$3.20	\$10.00	\$6.50	-	-
(Pd 2) Amount Spent	\$64.00	\$10.00	\$32.50	\$106.50	6.5%
(Pd 3) Price	\$3.10	\$10.00	\$7.00	-	-
(Pd 3) Amount Spent	\$62.00	\$10.00	\$35.00	\$107.00	0.5%
(Pd 4) Price	\$3.50	\$10.00	\$7.50	-	-
(Pd 4) Amount Spent	\$70.00	\$10.00	\$37.50	\$117.50	9.8%

TABLE 22.2 A College Student's Basket of Goods

To calculate the annual rate of inflation in this example:

Step 1. Find the percentage change in the cost of purchasing the overall basket of goods between the time periods. The general equation for percentage changes between two years, whether in the context of inflation or in any other calculation, is:

$$\frac{(\text{Level in new year} - \text{Level in previous year})}{\text{Level in previous year}} \times 100 = \text{Percentage change}$$

Step 2. From period 1 to period 2, the total cost of purchasing the basket of goods in [Table 22.2](#) rises from \$100 to \$106.50. Therefore, the percentage change over this time—the inflation rate—is:

$$\frac{(106.50 - 100)}{100.0} = 0.065 = 6.5\%$$

Step 3. From period 2 to period 3, the overall change in the cost of purchasing the basket rises from \$106.50 to \$107. Thus, the inflation rate over this time, again calculated by the percentage change, is approximately:

$$\frac{(107 - 106.50)}{106.50} = 0.0047 = 0.47\%$$

Step 4. From period 3 to period 4, the overall cost rises from \$107 to \$117.50. The inflation rate is thus:

$$\frac{(117.50 - 107)}{107} = 0.098 = 9.8\%$$

This calculation of the change in the total cost of purchasing a basket of goods accounts for how much a student spends on each good. Hamburgers are the lowest-priced good in this example, and aspirin is the highest-priced. If an individual buys a greater quantity of a low-price good, then it makes sense that changes in the price of that good should have a larger impact on the buying power of that person's money. The larger impact of hamburgers shows up in the "amount spent" row, where, in all time periods, hamburgers are the largest item within the amount spent row.

Index Numbers

The numerical results of a calculation based on a basket of goods can get a little messy. The simplified example in [Table 22.2](#) has only three goods and the prices are in even dollars, not numbers like 79 cents or \$124.99. If the list of products were much longer, and we used more realistic prices, the total quantity spent over a year might be some messy-looking number like \$17,147.51 or \$27,654.92.

To simplify the task of interpreting the price levels for more realistic and complex baskets of goods, economists typically report the price level in each period as an **index number**, rather than as the dollar amount for buying the basket of goods. Economists create price indices to calculate an overall average change in relative prices over time. To convert the money spent on the basket to an index number, economists arbitrarily choose one year to be the **base year**, or starting point from which we measure changes in prices. The base year, by definition, has an index number equal to 100. This sounds complicated, but it is really a simple math trick. In the example above, say that we choose time period 3 as the base year. Since the total amount of spending in that year is \$107, we divide that amount by itself (\$107) and multiply by 100. Again, this is because the index number in the base year *always* has to have a value of 100. Then, to figure out the values of the index number for the other years, we divide the dollar amounts for the other years by 1.07 as well. Note also that the dollar signs cancel out so that index numbers have no units.

[Table 22.3](#) shows calculations for the other values of the index number, based on the example in [Table 22.2](#). Because we calculate the index numbers so that they are in exactly the same proportion as the total dollar cost of purchasing the basket of goods, we can calculate the inflation rate based on the index numbers, using the percentage change formula. Thus, the inflation rate from period 1 to period 2 would be

$$\frac{(99.5 - 93.4)}{93.4} = 0.065 = 6.5\%$$

This is the same answer that we derived when measuring inflation based on the dollar cost of the basket of goods for the same time period.

	Total Spending	Index Number	Inflation Rate Since Previous Period
Period 1	\$100	$\frac{100}{1.07} = 93.4$	
Period 2	\$106.50	$\frac{106.50}{1.07} = 99.5$	$\frac{(99.5 - 93.4)}{93.4} = 0.065 = 6.5\%$
Period 3	\$107	$\frac{107}{1.07} = 100.0$	$\frac{100 - 99.5}{99.5} = 0.005 = 0.5\%$
Period 4	\$117.50	$\frac{117.50}{1.07} = 109.8$	$\frac{109.8 - 100}{100} = 0.098 = 9.8\%$

TABLE 22.3 Calculating Index Numbers When Period 3 is the Base Year

If the inflation rate is the same whether it is based on dollar values or index numbers, then why bother with the index numbers? The advantage is that indexing allows easier eyeballing of the inflation numbers. If you glance at two index numbers like 107 and 110, you know automatically that the rate of inflation between the two years is about, but not quite exactly equal to, 3%. By contrast, imagine that we express the price levels in absolute dollars of a large basket of goods, so that when you looked at the data, the numbers were \$19,493.62 and \$20,040.17. Most people find it difficult to eyeball those kinds of numbers and say that it is a change of about 3%. However, the two numbers expressed in absolute dollars are exactly in the same proportion of 107 to 110 as the previous example. If you're wondering why simple subtraction of the index numbers wouldn't work, read the following Clear It Up feature.



CLEAR IT UP

Why do you not just subtract index numbers?

A word of warning: When a price index moves from, say, 107 to 110, the rate of inflation is not *exactly* 3%. Remember, the inflation rate is not derived by subtracting the index numbers, but rather through the percentage-change calculation. We calculate the precise inflation rate as the price index moves from 107 to 110 as $100 \times (110 - 107) / 107 = 100 \times 0.028 = 2.8\%$. When the base year is fairly close to 100, a quick subtraction is not a terrible shortcut to calculating the inflation rate—but when precision matters down to tenths of a percent, subtracting will not give the right answer.

Two final points about index numbers are worth remembering. First, index numbers have no dollar signs or other units attached to them. Although we can use index numbers to calculate a percentage inflation rate, the index numbers themselves do not have percentage signs. Index numbers just mirror the proportions that we find in other data. They transform the other data so that it is easier to work with the data.

Second, the choice of a base year for the index number—that is, the year that is automatically set equal to 100—is arbitrary. We choose it as a starting point from which we can track changes in prices. In the official inflation statistics, it is common to use one base year for a few years, and then to update it, so that the base year of 100 is relatively close to the present. However, any base year that we choose for the index numbers will result in exactly the same inflation rate. To see this in the previous example, imagine that period 1 is the base year when total spending was \$100, and we assign it an index number of 100. At a glance, you can see that the index numbers would now exactly match the dollar figures, and the inflation rate in the first period would be 6.5%.

Now that we see how indexes work to track inflation, the next module will show us how economists measure the cost of living.



LINK IT UP

Watch this [video \(http://openstax.org/l/Duck_Tales\)](http://openstax.org/l/Duck_Tales) from the cartoon *Duck Tales* to view a mini-lesson on inflation.

22.2 How to Measure Changes in the Cost of Living

LEARNING OBJECTIVES

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

- Use the Consumer Price Index (CPI) to calculate U.S. inflation rates
- Identify several ways the Bureau of Labor Statistics avoids biases in the Consumer Price Index (CPI)
- Differentiate among the Consumer Price Index (CPI), the Producer Price Index (PPI), the International Price Index, the Employment Cost Index, and the GDP deflator.

The most commonly cited measure of inflation in the United States is the **Consumer Price Index (CPI)**.

Government statisticians at the U.S. Bureau of Labor Statistics calculate the CPI based on the prices in a fixed basket of goods and services that represents the purchases of the average family of four. In recent years, the statisticians have paid considerable attention to a subtle problem: that the change in the total cost of buying a fixed basket of goods and services over time is conceptually not quite the same as the change in the cost of living, because the cost of living represents how much it costs for a person to feel that their consumption provides an equal level of satisfaction or utility.

To understand the distinction, imagine that over the past 10 years, the cost of purchasing a fixed basket of goods increased by 25% and your salary also increased by 25%. Has your personal standard of living held constant? If you do not necessarily purchase an identical fixed basket of goods every year, then an inflation calculation based on the cost of a fixed basket of goods may be a misleading measure of how your cost of living has changed. Two problems arise here: substitution bias and quality/new goods bias.

When the price of a good rises, consumers tend to purchase less of it and to seek out substitutes instead. Conversely, as the price of a good falls, people will tend to purchase more of it. This pattern implies that goods with generally rising prices should tend over time to become less important in the overall basket of goods used to calculate inflation, while goods with falling prices should tend to become more important. Consider, as an example, a rise in the price of peaches by \$100 per pound. If consumers were utterly inflexible in their demand for peaches, this would lead to a big rise in the price of food for consumers. Alternatively, imagine that people are utterly indifferent to whether they have peaches or other types of fruit. Now, if peach prices rise, people completely switch to other fruit choices and the average price of food does not change at all. A fixed and unchanging basket of goods assumes that consumers are locked into buying exactly the same goods, regardless of price changes—not a very likely assumption. Thus, **substitution bias**—the rise in the price of a fixed basket of goods over time—tends to overstate the rise in a consumer's true cost of living, because it does not take into account that the person can substitute away from goods whose relative prices have risen.

The other major problem in using a fixed basket of goods as the basis for calculating inflation is how to deal with the arrival of improved versions of older goods or altogether new goods. Consider the problem that arises if a cereal is improved by adding 12 essential vitamins and minerals—and also if a box of the cereal costs 5% more. It would clearly be misleading to count the entire resulting higher price as inflation, because the new price reflects a higher quality (or at least different) product. Ideally, one would like to know how much of the higher price is due to the quality change, and how much of it is just a higher price. The Bureau of Labor Statistics, which is responsible for computing the Consumer Price Index, must deal with these difficulties in adjusting for quality changes.

LINK IT UP

Visit this [website \(http://openstax.org/l/Fords\)](http://openstax.org/l/Fords) to view a list of Ford car prices between 1909 and 1927. Consider how these prices compare to today's models. Is the product today of a different quality?

We can think of a new product as an extreme improvement in quality—from something that did not exist to something that does. However, the basket of goods that was fixed in the past obviously does not include new goods created since then. The basket of goods and services in the Consumer Price Index (CPI) is revised and updated over time, and so new products are gradually included. However, the process takes some time. For example, room air conditioners were widely sold in the early 1950s, but were not introduced into the basket of goods behind the Consumer Price Index until 1964. The VCR and personal computer were available in the late 1970s and widely sold by the early 1980s, but did not enter the CPI basket of goods until 1987. By 1996, there were more than 40 million cellular phone subscribers in the United States—but cell phones were not yet part of the CPI basket of goods. The parade of inventions has continued, with the CPI inevitably lagging a few years behind.

The arrival of new goods creates problems with respect to the accuracy of measuring inflation. The reason

people buy new goods, presumably, is that the new goods offer better value for money than existing goods. Thus, if the price index leaves out new goods, it overlooks one of the ways in which the cost of living is improving. In addition, the price of a new good is often higher when it is first introduced and then declines over time. If the new good is not included in the CPI for some years, until its price is already lower, the CPI may miss counting this price decline altogether. Taking these arguments together, the **quality/new goods bias** means that the rise in the price of a fixed basket of goods over time tends to overstate the rise in a consumer's true cost of living, because it does not account for how improvements in the quality of existing goods or the invention of new goods improves the standard of living. The following Clear It Up feature is a must-read on how statisticians comprise and calculate the CPI.



CLEAR IT UP

How do U.S. government statisticians measure the Consumer Price Index?

When the U.S. Bureau of Labor Statistics (BLS) calculates the Consumer Price Index, the first task is to decide on a basket of goods that is representative of the purchases of the average household. We do this by using the Consumer Expenditure Survey, a national survey of about 7,000 households, which provides detailed information on spending habits. Statisticians divide consumer expenditures into eight major groups (seen below), which in turn they divide into more than 200 individual item categories. The BLS currently uses 1982–1984 as the base period.

For each of the 200 individual expenditure items, the BLS chooses several hundred very specific examples of that item and looks at the prices of those examples. In figuring out the “breakfast cereal” item under the overall category of “foods and beverages,” the BLS picks several hundred examples of breakfast cereal. One example might be the price of a 24-oz. box of a particular brand of cereal sold at a particular store. The BLS statistically selects specific products and sizes and stores to reflect what people buy and where they shop. The basket of goods in the Consumer Price Index thus consists of about 80,000 products; that is, several hundred specific products in over 200 broad-item categories. Statisticians rotate about one-quarter of these 80,000 specific products of the sample each year, and replace them with a different set of products.

The next step is to collect data on prices. Data collectors visit or call about 23,000 stores in 87 urban areas all over the United States every month to collect prices on these 80,000 specific products. The BLS also conducts a survey of 50,000 landlords or tenants to collect information about rents.

Statisticians then calculate the Consumer Price Index by taking the 80,000 prices of individual products and combining them, using weights (see [Figure 22.2](#)) determined by the quantities of these products that people buy and allowing for factors like substitution between goods and quality improvements, into price indices for the 200 or so overall items. Then, the statisticians combine the price indices for the 200 items into an overall Consumer Price Index. According to the Consumer Price Index website, there are eight categories that data collectors use:

The Eight Major Categories in the Consumer Price Index

1. Food and beverages (breakfast cereal, milk, coffee, chicken, wine, full-service meals, and snacks)
2. Housing (renter's cost of housing, homeowner's cost of housing, fuel oil, bedroom furniture)
3. Apparel (men's shirts and sweaters, women's dresses, jewelry)
4. Transportation (new vehicles, airline fares, gasoline, motor vehicle insurance)
5. Medical care (prescription drugs and medical supplies, physicians' services, eyeglasses and eye care, hospital services)
6. Recreation (televisions, cable television, pets and pet products, sports equipment, admissions)
7. Education and communication (college tuition, postage, telephone services, computer software and accessories)
8. Other goods and services (tobacco and smoking products, haircuts and other personal services, funeral expenses)

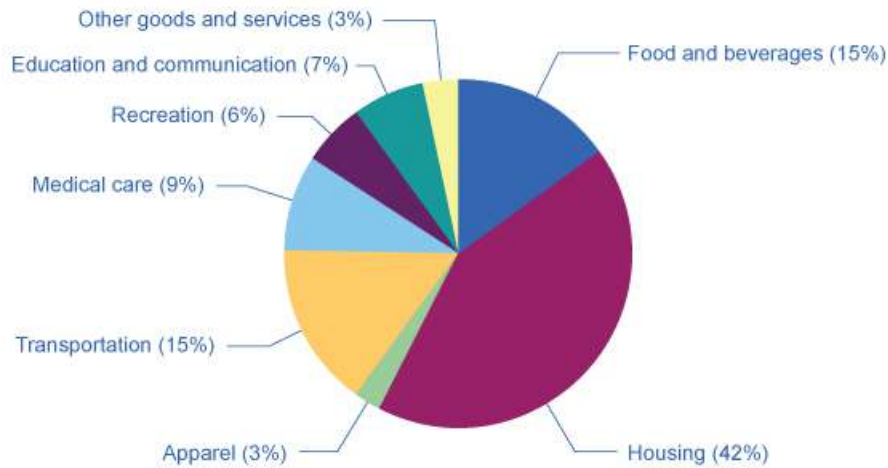


FIGURE 22.2 The Weighting of CPI Components Of the eight categories used to generate the Consumer Price Index, housing is the highest at 42.4%. The next highest category, food and beverage at 15.1%, is less than half the size of housing. Other goods and services, and apparel, are the lowest at 3.2% and 2.7%, respectively. (Source: www.bls.gov/cpi)

The CPI and Core Inflation Index

Imagine if you were driving a company truck across the country- you probably would care about things like the prices of available roadside food and motel rooms as well as the truck's operating condition. However, the manager of the firm might have different priorities. He would care mostly about the truck's on-time performance and much less so about the food you were eating and the places you were staying. In other words, the company manager would be paying attention to the firm's production, while ignoring transitory elements that impacted you, but did not affect the company's bottom line.

In a sense, a similar situation occurs with regard to measures of inflation. As we've learned, CPI measures prices as they affect everyday household spending. Economists typically calculate a **core inflation index** by taking the CPI and excluding volatile economic variables. In this way, economists have a better sense of the underlying trends in prices that affect the cost of living.

Examples of excluded variables include energy and food prices, which can jump around from month to month because of the weather. According to an article by Kent Bernhard, during Hurricane Katrina in 2005, a key supply point for the nation's gasoline was nearly knocked out. Gas prices quickly shot up across the nation, in some places by up to 40 cents a gallon in one day. This was not the cause of an economic policy but rather a short-lived event until the pumps were restored in the region. In this case, the CPI that month would register the change as a cost of living event to households, but the core inflation index would remain unchanged. As a result, the Federal Reserve's decisions on interest rates would not be influenced. Similarly, droughts can cause world-wide spikes in food prices that, if temporary, do not affect the nation's economic capability.

As former Chairman of the Federal Reserve Ben Bernanke noted in 1999 about the core inflation index, "It provide(s) a better guide to monetary policy than the other indices, since it measures the more persistent underlying inflation rather than transitory influences on the price level." Bernanke also noted that it helps communicate that the Federal Reserve does not need to respond to every inflationary shock since some price changes are transitory and not part of a structural change in the economy.

In sum, both the CPI and the core inflation index are important, but serve different audiences. The CPI helps households understand their overall cost of living from month to month, while the core inflation index is a preferred gauge from which to make important government policy changes.

Practical Solutions for the Substitution and the Quality/New Goods Biases

By the early 2000s, the Bureau of Labor Statistics was using alternative mathematical methods for calculating the Consumer Price Index, more complicated than just adding up the cost of a fixed basket of goods, to allow for some substitution between goods. It was also updating the basket of goods behind the CPI more frequently, so that it could include new and improved goods more rapidly. For certain products, the BLS was carrying out studies to try to measure the quality improvement. For example, with computers, an economic study can try to adjust for changes in speed, memory, screen size, and other product characteristics, and then calculate the change in price after accounting for these product changes. However, these adjustments are inevitably imperfect, and exactly how to make these adjustments is often a source of controversy among professional economists.

By the early 2000s, the substitution bias and quality/new goods bias had been somewhat reduced, so that since then the rise in the CPI probably overstates the true rise in inflation by only about 0.5% per year. Over one or a few years, this is not much. Over a period of a decade or two, even half of a percent per year compounds to a more significant amount. In addition, the CPI tracks prices from physical locations, and not at online sites like Amazon, where prices can be lower.

When measuring inflation (and other economic statistics, too), a tradeoff arises between simplicity and interpretation. If we calculate the inflation rate with a basket of goods that is fixed and unchanging, then the calculation of an inflation rate is straightforward, but the problems of substitution bias and quality/new goods bias will arise. However, when the basket of goods is allowed to shift and evolve to reflect substitution toward lower relative prices, quality improvements, and new goods, the technical details of calculating the inflation rate grow more complex.

Additional Price Indices: PPI, GDP Deflator, and More

The basket of goods behind the Consumer Price Index represents an average hypothetical U.S. household's consumption, which is to say that it does not exactly capture anyone's personal experience. When the task is to calculate an average level of inflation, this approach works fine. What if, however, you are concerned about inflation experienced by a certain group, like the elderly, or the poor, or single-parent families with children, or Hispanic-Americans? In specific situations, a price index based on the buying power of the average consumer may not feel quite right.

This problem has a straightforward solution. If the Consumer Price Index does not serve the desired purpose, then invent another index, based on a basket of goods appropriate for the group of interest. The Bureau of Labor Statistics publishes a number of experimental price indices: some for particular groups like the elderly or the poor, some for different geographic areas, and some for certain broad categories of goods like food or housing.

The BLS also calculates several price indices that are not based on baskets of consumer goods. For example, the **Producer Price Index (PPI)** is based on prices paid for supplies and inputs by producers of goods and services. We can break it down into price indices for different industries, commodities, and stages of processing (like finished goods, intermediate goods, or crude materials for further processing). There is an **International Price Index** based on the prices of merchandise that is exported or imported. An **Employment Cost Index** measures wage inflation in the labor market. The **GDP deflator**, which the Bureau of Economic Analysis measures, is a price index that includes all the GDP components (that is, consumption plus investment plus government plus exports minus imports). Unlike the CPI, its baskets are not fixed but recalculate what that year's GDP would have been worth using the base-year's prices. MIT's Billion Prices Project is a more recent alternative attempt to measure prices: economists collect data online from retailers and then put them into an index that they compare to the CPI (Source: <http://bpp.mit.edu/usa/>).

What's the best measure of inflation? If one is concerned with the most accurate measure of inflation, one should use the GDP deflator as it picks up the prices of goods and services produced. However, it is not a good

measure of the cost of living as it includes prices of many products not purchased by households (for example, aircraft, fire engines, factory buildings, office complexes, and bulldozers). If one wants the most accurate measure of inflation as it impacts households, one should use the CPI, as it only picks up prices of products purchased by households. That is why economists sometimes refer to the CPI as the cost-of-living index. As the Bureau of Labor Statistics states on its website: “The ‘best’ measure of inflation for a given application depends on the intended use of the data.”

22.3 How the U.S. and Other Countries Experience Inflation

LEARNING OBJECTIVES

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

- Identify patterns of inflation for the United States using data from the Consumer Price Index
- Identify patterns of inflation on an international level

In the last three decades, inflation has been relatively low in the U.S. economy, with the Consumer Price Index typically rising 2% to 4% per year. Looking back over the twentieth century, there have been several periods where inflation caused the price level to rise at double-digit rates, but nothing has come close to hyperinflation.

Historical Inflation in the U.S. Economy

Figure 22.3 (a) shows the level of prices in the Consumer Price Index stretching back to 1913. In this case, the base years (when the CPI is defined as 100) are set for the average level of prices that existed from 1982 to 1984. Figure 22.3 (b) shows the annual percentage changes in the CPI over time, which is the inflation rate.

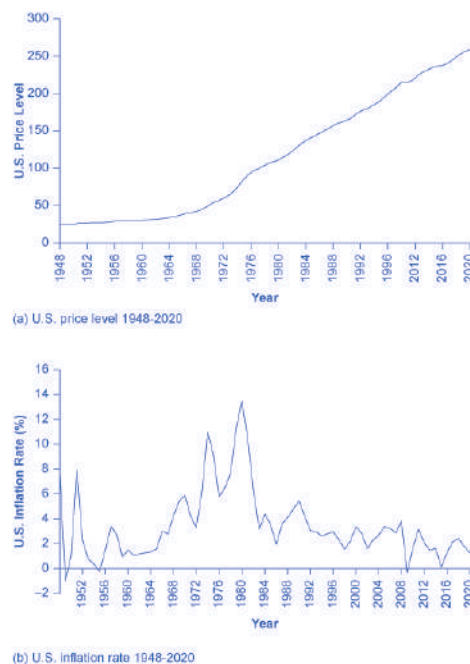


FIGURE 22.3 U.S. Price Level and Inflation Rates since 1947 Graph a shows the trends in the U.S. price level from the year 1947 to 2020. In 1947, the graph starts out close to 22. It gradually increases until about 1973, then increases more rapidly through the remainder of the 1970s and beyond, with periodic dips, until 2020, when it reached around 260. Graph b shows the trends in U.S. inflation rates from the year 1948 to 2020. In 1948, the graph starts out with inflation at almost 7%, goes up and down periodically, with peaks in the 1940s and the 1970s, until settling to around 1.2% in 2020.

[Click to view content \(https://openstax.org/books/principles-economics-3e/pages/22-3-how-the-us-and-other-countries-experience-inflation\)](https://openstax.org/books/principles-economics-3e/pages/22-3-how-the-us-and-other-countries-experience-inflation)

Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly.

The first two waves of inflation are easy to characterize in historical terms: they are right after World War I and World War II. However, there are also two periods of severe negative inflation—called **deflation**—in the early decades of the twentieth century: one following the deep 1920–21 recession and the other during the Great Depression of the 1930s. (Since inflation is a time when the buying power of money in terms of goods and services is reduced, deflation will be a time when the buying power of money in terms of goods and services increases.) For the period from 1900 to about 1960, the major inflations and deflations nearly balanced each other out, so the average annual rate of inflation over these years was only about 1% per year. A third wave of more severe inflation arrived in the 1970s and departed in the early 1980s.

LINK IT UP

Visit this [website \(http://openstax.org/l/CPI_calculator\)](http://openstax.org/l/CPI_calculator) to use an inflation calculator and discover how prices have changed in the last 100 years.

Times of recession or depression often seem to be times when the inflation rate is lower, as in the recession of 1920–1921, the Great Depression, the recession of 1980–1982, and the Great Recession in 2008–2009. There were a few months in 2009 that were deflationary, but not at an annual rate. High levels of unemployment typically accompany recessions, and the total demand for goods falls, pulling the price level down. Conversely, the rate of inflation often, but not always, seems to start moving up when the economy is growing very strongly, like right after wartime or during the 1960s. The frameworks for macroeconomic analysis, that we developed in other chapters, will explain why recession often accompanies higher unemployment and lower inflation, while rapid economic growth often brings lower unemployment but higher inflation.

Inflation around the World

Around the rest of the world, the pattern of inflation has been very mixed; [Figure 22.4](#) shows inflation rates over the last several decades. Many industrialized countries, not just the United States, had relatively high inflation rates in the 1970s. For example, in 1975, Japan's inflation rate was over 8% and the inflation rate for the United Kingdom was almost 25%. In the 1980s, inflation rates came down in the United States and in Europe and have largely stayed down.

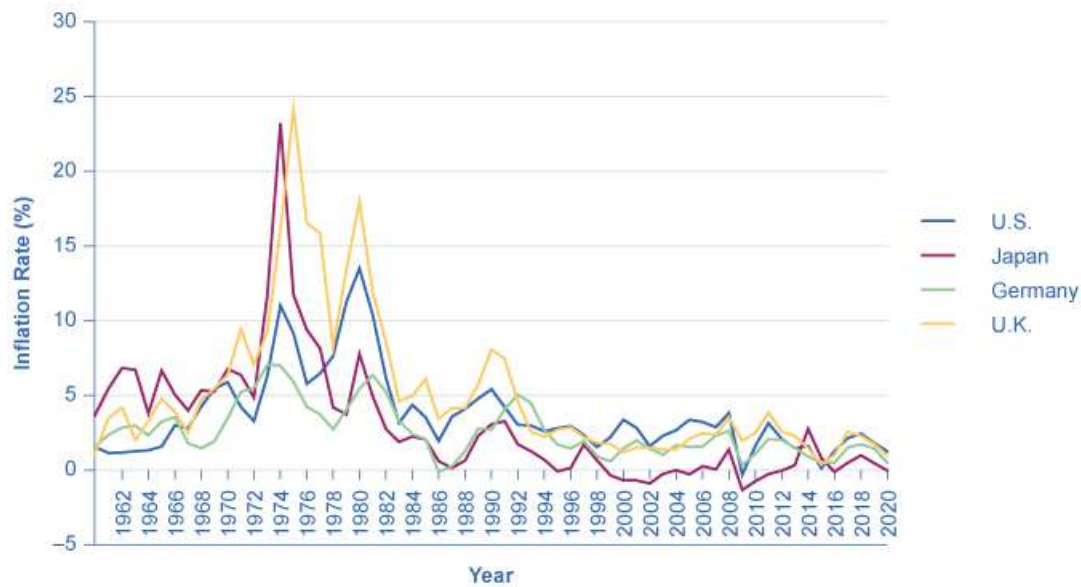
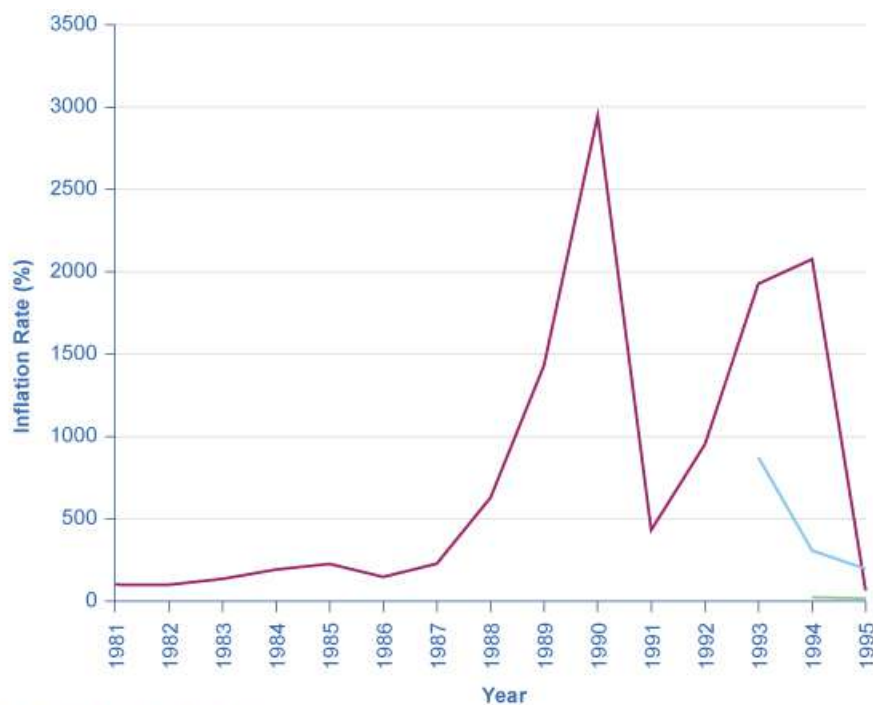
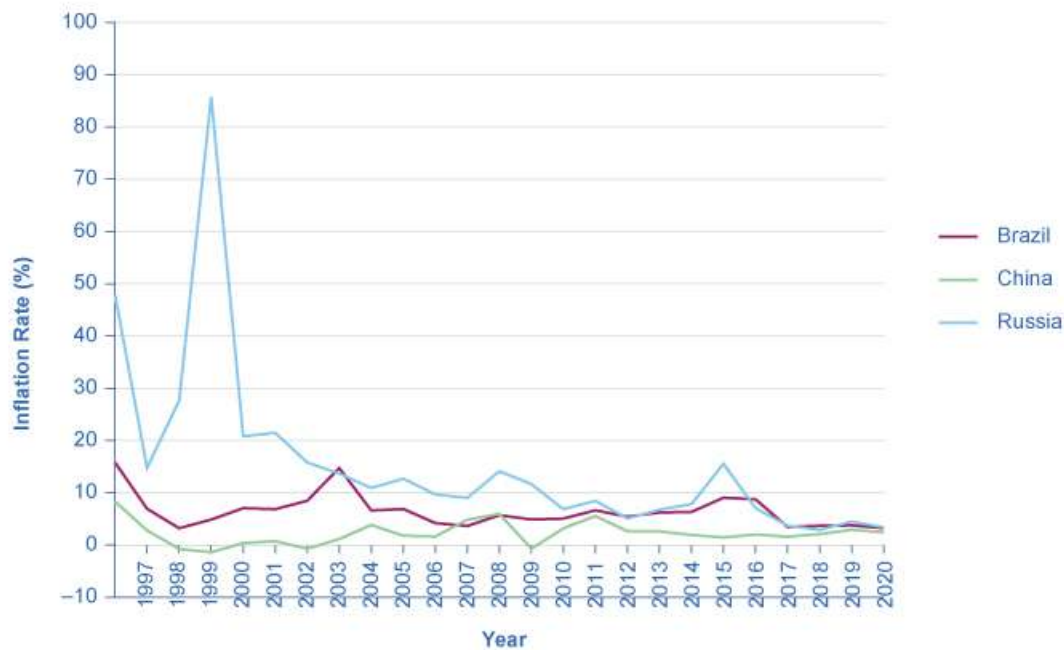


FIGURE 22.4 Countries with Relatively Low Inflation Rates, 1961–2020 This chart shows the annual percentage change in consumer prices compared with the previous year’s consumer prices in the United States, the United Kingdom, Japan, and Germany.

Countries with controlled economies in the 1970s, like the Soviet Union and China, historically had very low rates of measured inflation—because prices were forbidden to rise by law, except for the cases where the government deemed a price increase to be due to quality improvements. However, these countries also had perpetual shortages of goods, since forbidding prices to rise acts like a price ceiling and creates a situation where quantity demanded often exceeds quantity supplied. As Russia and China made a transition toward more market-oriented economies, they also experienced outbursts of inflation, although we should regard the statistics for these economies as somewhat shakier. Inflation in China averaged about 10% per year for much of the 1980s and early 1990s, although it has dropped off since then. Russia experienced **hyperinflation**—an outburst of high inflation—of 2,500% per year in the early 1990s, although by 2006 Russia’s consumer price inflation had dipped below 10% per year, as [Figure 22.5](#) shows. The closest the United States has ever reached hyperinflation was during the 1860–1865 Civil War, in the Confederate states.



(a) Inflation rates 1981-1995



(b) Inflation rates 1996-2020

FIGURE 22.5 Countries with Relatively High Inflation Rates, 1981–2020 These charts show the percentage change in consumer prices compared with the previous year’s consumer prices in Brazil, China, and Russia. (a) Of these, Brazil and Russia experienced very high inflation at some point between the late-1980s and late-1990s. (b) Though not as high, China also had high inflation rates in the mid-1990s. Even though their inflation rates have come down over the last two decades, several of these countries continue to see significant inflation rates. (Sources: <http://www.inflation.eu/inflation-rates>; <http://research.stlouisfed.org/fred2/series/FPCPITOTLZGBRA>; <http://research.stlouisfed.org/fred2/series/CHNCPIALLMINMEI>; <http://research.stlouisfed.org/fred2/series/FPCPITOTLZGRUS>)

Many countries in Latin America experienced raging inflation during the 1980s and early 1990s, with inflation rates often well above 100% per year. In 1990, for example, both Brazil and Argentina saw inflation climb above 2000%. Certain countries in Africa experienced extremely high rates of inflation, sometimes bordering on hyperinflation, in the 1990s. Nigeria, the most populous country in Africa, had an inflation rate of 75% in 1995.

In the early 2000s, the problem of inflation appears to have diminished for most countries, at least in comparison to the worst times of recent decades. As we noted in this earlier Bring it Home feature, in recent years, the world's worst example of hyperinflation was in Zimbabwe, where at one point the government was issuing bills with a face value of \$100 trillion (in Zimbabwean dollars)—that is, the bills had \$100,000,000,000,000 written on the front, but were almost worthless. In many countries, the memory of double-digit, triple-digit, and even quadruple-digit inflation is not very far in the past.

22.4 The Confusion Over Inflation

LEARNING OBJECTIVES

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

- Explain how inflation can cause redistributions of purchasing power
- Identify ways inflation can blur the perception of supply and demand
- Explain the economic benefits and challenges of inflation

Economists usually oppose high inflation, but they oppose it in a milder way than many non-economists. Robert Shiller, one of 2013's Nobel Prize winners in economics, carried out several surveys during the 1990s about attitudes toward inflation. One of his questions asked, "Do you agree that preventing high inflation is an important national priority, as important as preventing drug use or preventing deterioration in the quality of our schools?" Answers were on a scale of 1–5, where 1 meant "Fully agree" and 5 meant "Completely disagree." For the U.S. population as a whole, 52% answered "Fully agree" that preventing high inflation was a highly important national priority and just 4% said "Completely disagree." However, among professional economists, only 18% answered "Fully agree," while the same percentage of 18% answered "Completely disagree."

The Land of Funny Money

What are the economic problems caused by inflation, and why do economists often regard them with less concern than the general public? Consider a very short story: "The Land of Funny Money."

One morning, everyone in the Land of Funny Money awakened to find that everything denominated in money had increased by 20%. The change was completely unexpected. Every price in every store was 20% higher. Paychecks were 20% higher. Interest rates were 20 % higher. The amount of money, everywhere from wallets to savings accounts, was 20% larger. This overnight inflation of prices made newspaper headlines everywhere in the Land of Funny Money. However, the headlines quickly disappeared, as people realized that in terms of what they could actually buy with their incomes, this inflation had no economic impact. Everyone's pay could still buy exactly the same set of goods as it did before. Everyone's savings were still sufficient to buy exactly the same car, vacation, or retirement that they could have bought before. Equal levels of inflation in all wages and prices ended up not mattering much at all.

When the people in Robert Shiller's surveys explained their concern about inflation, one typical reason was that they feared that as prices rose, they would not be able to afford to buy as much. In other words, people were worried because they did not live in a place like the Land of Funny Money, where all prices and wages rose simultaneously. Instead, people live here on Planet Earth, where prices might rise while wages do not rise at all, or where wages rise more slowly than prices.

Economists note that over most periods, the inflation level in prices is roughly similar to the inflation level in wages, and so they reason that, on average, over time, people's economic status is not greatly changed by

inflation. If all prices, wages, and interest rates adjusted automatically and immediately with inflation, as in the Land of Funny Money, then no one's purchasing power, profits, or real loan payments would change. However, if other economic variables do not move exactly in sync with inflation, or if they adjust for inflation only after a time lag, then inflation can cause three types of problems: unintended redistributions of purchasing power, blurred price signals, and difficulties in long-term planning.

Unintended Redistributions of Purchasing Power

Inflation can cause redistributions of purchasing power that hurt some and help others. People who are hurt by inflation include those who are holding considerable cash, whether it is in a safe deposit box or in a cardboard box under the bed. When inflation happens, the buying power of cash diminishes. However, cash is only an example of a more general problem: anyone who has financial assets invested in a way that the nominal return does not keep up with inflation will tend to be affected by inflation. For example, if a person has money in a bank account that pays 4% interest, but inflation rises to 5%, then the real rate of return for the money invested in that bank account is negative 1%.

The problem of a good-looking nominal interest rate transforming into an ugly-looking real interest rate can be worsened by taxes. The U.S. income tax is charged on the nominal interest received in dollar terms, without an adjustment for inflation. Thus, the government taxes a person who invests \$10,000 and receives a 5% nominal rate of interest on the \$500 received—no matter whether the inflation rate is 0%, 5%, or 10%. If inflation is 0%, then the real interest rate is 5% and all \$500 is a gain in buying power. However, if inflation is 5%, then the real interest rate is zero and the person had no real gain—but owes income tax on the nominal gain anyway. If inflation is 10%, then the real interest rate is *negative* 5% and the person is actually falling behind in buying power, but would still owe taxes on the \$500 in nominal gains.

Inflation can cause unintended redistributions for wage earners, too. Wages do typically creep up with inflation over time, eventually. The last row of [Table 22.1](#) at the start of this chapter showed that the average hourly wage in manufacturing in the U.S. economy increased from \$3.23 in 1970 to \$30.11 in 2021, which is an increase by a factor of close to ten. Over that time period, the Consumer Price Index increased by about a factor of eight. However, increases in wages may lag behind inflation for a year or two, since wage adjustments are often somewhat sticky and occur only once or twice a year. Moreover, the extent to which wages keep up with inflation creates insecurity for workers and may involve painful, prolonged conflicts between employers and employees. If the government adjusts minimum wage for inflation only infrequently, minimum wage workers are losing purchasing power from their nominal wages, as [Figure 22.6](#) shows.

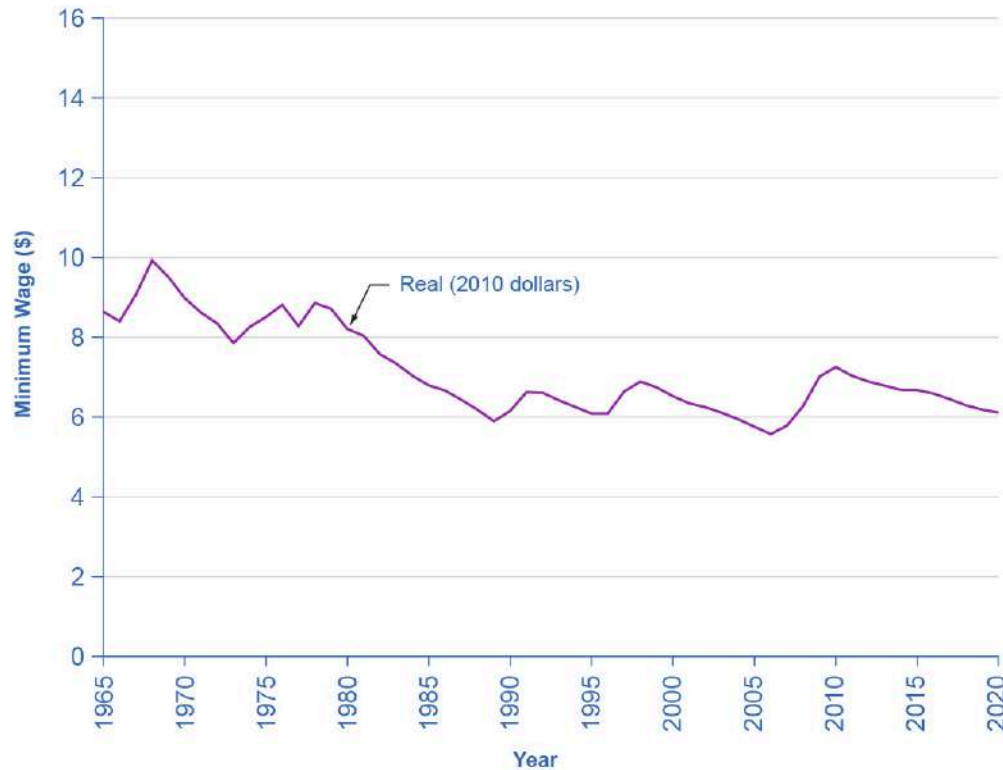


FIGURE 22.6 U.S. Minimum Wage and Inflation After adjusting for inflation, the federal minimum wage dropped about 30 percent from 1965–2020, even though the nominal figure climbed from \$1.40 to \$7.25 per hour. Increases in the minimum wage in between 2008 and 2010 kept the decline from being worse—as it would have been if the wage had remained the same as it did from 1997 through 2007. Since 2010, the real minimum wage has continued to decline. (Sources: <http://www.dol.gov/whd/minwage/chart.htm> and <http://data.bls.gov/cgi-bin/surveymost?cu>)

One sizable group of people has often received a large share of their income in a form that does not increase over time: retirees who receive a private company pension. Most pensions have traditionally been set as a fixed nominal dollar amount per year at retirement. For this reason, economists call pensions “defined benefits” plans. Even if inflation is low, the combination of inflation and a fixed income can create a substantial problem over time. A person who retires on a fixed income at age 65 will find that losing just 1% to 2% of buying power per year to inflation compounds to a considerable loss of buying power after a decade or two.

Fortunately, pensions and other defined benefits retirement plans are increasingly rare, replaced instead by “defined contribution” plans, such as 401(k)s and 403(b)s. In these plans, the employer contributes a fixed amount to the worker’s retirement account on a regular basis (usually every pay check). The employee often contributes as well. The worker invests these funds in a wide range of investment vehicles. These plans are tax deferred, and they are portable so that if the individual takes a job with a different employer, their 401(k) comes with them. To the extent that the investments made generate real rates of return, retirees are not burdened by the inflation costs of traditional pensioners.

However, ordinary people can sometimes benefit from the unintended redistributions of inflation. Consider someone who borrows \$10,000 to buy a car at a fixed interest rate of 9%. If inflation is 3% at the time the loan is made, then the borrower must repay the loan at a real interest rate of 6%. However, if inflation rises to 9%, then the real interest rate on the loan is zero. In this case, the borrower’s benefit from inflation is the lender’s loss. A borrower paying a fixed interest rate, who benefits from inflation, is just the flip side of an investor receiving a fixed interest rate, who suffers from inflation. The lesson is that when interest rates are fixed, rises in the rate of inflation tend to penalize suppliers of financial capital, who receive repayment in dollars that are

worth less because of inflation, while demanders of financial capital end up better off, because they can repay their loans in dollars that are worth less than originally expected.

The unintended redistributions of buying power that inflation causes may have a broader effect on society. America's widespread acceptance of market forces rests on a perception that people's actions have a reasonable connection to market outcomes. When inflation causes a retiree who built up a pension or invested at a fixed interest rate to suffer, however, while someone who borrowed at a fixed interest rate benefits from inflation, it is hard to believe that this outcome was deserved in any way. Similarly, when homeowners benefit from inflation because the price of their homes rises, while renters suffer because they are paying higher rent, it is hard to see any useful incentive effects. One of the reasons that the general public dislikes inflation is a sense that it makes economic rewards and penalties more arbitrary—and therefore likely to be perceived as unfair – even dangerous, as the next Clear It Up feature shows.



CLEAR IT UP

Is there a connection between German hyperinflation and Hitler's rise to power?

Germany suffered an intense hyperinflation of its currency, the Mark, in the years after World War I, when the Weimar Republic in Germany resorted to printing money to pay its bills and the onset of the Great Depression created the social turmoil that Adolf Hitler was able to take advantage of in his rise to power. Shiller described the connection this way in a National Bureau of Economic Research 1996 Working Paper:

A fact that is probably little known to young people today, even in Germany, is that the final collapse of the Mark in 1923, the time when the Mark's inflation reached astronomical levels (inflation of 35,974.9% in November 1923 alone, for an annual rate that month of $4.69 \times 10^{28}\%$), came in the same month as did Hitler's Beer Hall Putsch, his Nazi Party's armed attempt to overthrow the German government. This failed putsch resulted in Hitler's imprisonment, at which time he wrote his book *Mein Kampf*, setting forth an inspirational plan for Germany's future, suggesting plans for world domination. . .

. . . Most people in Germany today probably do not clearly remember these events; this lack of attention to it may be because its memory is blurred by the more dramatic events that succeeded it (the Nazi seizure of power and World War II). However, to someone living through these historical events in sequence . . . [the putsch] may have been remembered as vivid evidence of the potential effects of inflation.

Blurred Price Signals

Prices are the messengers in a market economy, conveying information about conditions of demand and supply. Inflation blurs those price messages. Inflation means that we perceive price signals more vaguely, like a radio program received with considerable static. If the static becomes severe, it is hard to tell what is happening.

In Israel, when inflation accelerated to an annual rate of 500% in 1985, some stores stopped posting prices directly on items, since they would have had to put new labels on the items or shelves every few days to reflect inflation. Instead, a shopper just took items from a shelf and went up to the checkout register to find out the price for that day. Obviously, this situation makes comparing prices and shopping for the best deal rather difficult. When the levels and changes of prices become uncertain, businesses and individuals find it harder to react to economic signals. In a world where inflation is at a high rate, but bouncing up and down to some extent, does a higher price of a good mean that inflation has risen, or that supply of that good has decreased, or that demand for that good has increased? Should a buyer of the good take the higher prices as an economic hint to start substituting other products—or have the prices of the substitutes risen by an equal amount? Should a seller of the good take a higher price as a reason to increase production—or is the higher price only a sign of a general inflation in which the prices of all inputs to production are rising as well? The true story will presumably become clear over time, but at a given moment, who can say?

High and variable inflation means that the incentives in the economy to adjust in response to changes in prices are weaker. Markets will adjust toward their equilibrium prices and quantities more erratically and slowly, and many individual markets will experience a greater chance of surpluses and shortages.

Problems of Long-Term Planning

Inflation can make long-term planning difficult. In discussing unintended redistributions, we considered the case of someone trying to plan for retirement with a pension that is fixed in nominal terms and a high rate of inflation. Similar problems arise for all people trying to save for retirement, because they must consider what their money will really buy several decades in the future when we cannot know the rate of future inflation.

Inflation, especially at moderate or high levels, will pose substantial planning problems for businesses, too. A firm can make money from inflation—for example, by paying bills and wages as late as possible so that it can pay in inflated dollars, while collecting revenues as soon as possible. A firm can also suffer losses from inflation, as in the case of a retail business that gets stuck holding too much cash, only to see inflation eroding the value of that cash. However, when a business spends its time focusing on how to profit by inflation, or at least how to avoid suffering from it, an inevitable tradeoff strikes: less time is spent on improving products and services or on figuring out how to make existing products and services more cheaply. An economy with high inflation rewards businesses that have found clever ways of profiting from inflation, which are not necessarily the businesses that excel at productivity, innovation, or quality of service.

In the short term, low or moderate levels of inflation may not pose an overwhelming difficulty for business planning, because costs of doing business and sales revenues may rise at similar rates. If, however, inflation varies substantially over the short or medium term, then it may make sense for businesses to stick to shorter-term strategies. The evidence as to whether relatively low rates of inflation reduce productivity is controversial among economists. There is some evidence that if inflation can be held to moderate levels of less than 3% per year, it need not prevent a nation's real economy from growing at a healthy pace. For some countries that have experienced hyperinflation of several thousand percent per year, an annual inflation rate of 20–30% may feel basically the same as zero. However, several economists have pointed to the suggestive fact that when U.S. inflation heated up in the early 1970s—to 10%—U.S. growth in productivity slowed down, and when inflation slowed down in the 1980s, productivity edged up again not long thereafter, as [Figure 22.7](#) shows.

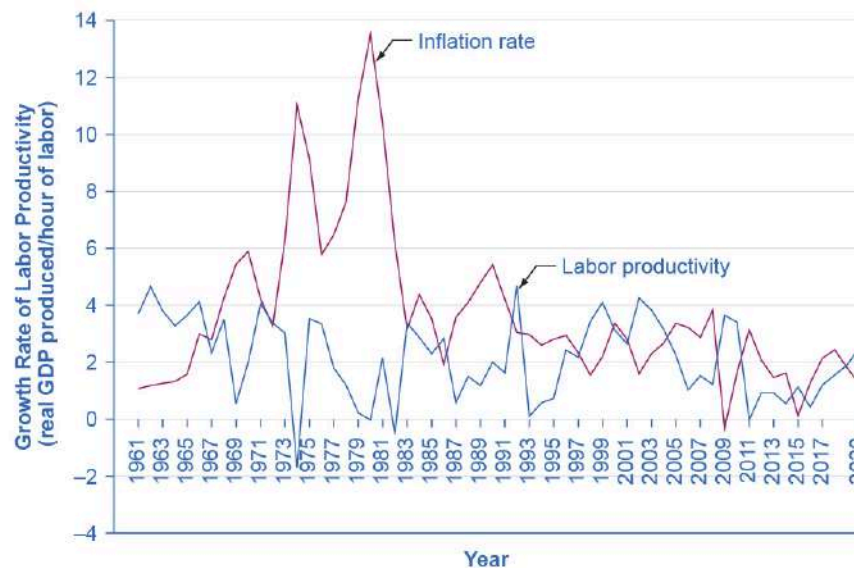


FIGURE 22.7 U.S. Inflation Rate and U.S. Labor Productivity, 1961–2020 Over the last several decades in the United States, there have been times when rising inflation rates have been closely followed by lower productivity rates and lower inflation rates have corresponded to increasing productivity rates. As the graph shows, however, this correlation does not always exist.

Any Benefits of Inflation?

Although the economic effects of inflation are primarily negative, two countervailing points are worth noting. First, the impact of inflation will differ considerably according to whether it is creeping up slowly at 0% to 2% per year, galloping along at 10% to 20% per year, or racing to the point of hyperinflation at, say, 40% per month. Hyperinflation can rip an economy and a society apart. An annual inflation rate of 2%, 3%, or 4%, however, is a long way from a national crisis. Low inflation is also better than deflation which occurs with severe recessions.

Second, economists sometimes argue that moderate inflation may help the economy by making wages in labor markets more flexible. The discussion in [Unemployment](#) pointed out that wages tend to be sticky in their downward movements and that unemployment can result. A little inflation could nibble away at real wages, and thus help real wages to decline if necessary. In this way, even if a moderate or high rate of inflation may act as sand in the gears of the economy, perhaps a low rate of inflation serves as oil for the gears of the labor market. This argument is controversial. A full analysis would have to account for all the effects of inflation. It does, however, offer another reason to believe that, all things considered, very low rates of inflation may not be especially harmful.

22.5 Indexing and Its Limitations

LEARNING OBJECTIVES

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

- Explain the relationship between indexing and inflation
- Identify three ways the government can control inflation through macroeconomic policy

When a price, wage, or interest rate is adjusted automatically with inflation, economists use the term **indexed**. An indexed payment increases according to the index number that measures inflation. Those in private markets and government programs observe a wide range of indexing arrangements. Since the negative effects of inflation depend in large part on having inflation unexpectedly affect one part of the economy but not another—say, increasing the prices that people pay but not the wages that workers receive—indexing will take some of the sting out of inflation.

Indexing in Private Markets

In the 1970s and 1980s, labor unions commonly negotiated wage contracts that had **cost-of-living adjustments (COLAs)** which guaranteed that their wages would keep up with inflation. These contracts were sometimes written as, for example, COLA plus 3%. Thus, if inflation was 5%, the wage increase would automatically be 8%, but if inflation rose to 9%, the wage increase would automatically be 12%. COLAs are a form of indexing applied to wages.

Loans often have built-in inflation adjustments, too, so that if the inflation rate rises by two percentage points, then the interest rate that a financial institution charges on the loan rises by two percentage points as well. An **adjustable-rate mortgage (ARM)** is a type of loan that one can use to purchase a home in which the interest rate varies with the rate of inflation. Often, a borrower will be able to receive a lower interest rate if borrowing with an ARM, compared to a fixed-rate loan. The reason is that with an ARM, the lender is protected against the risk that higher inflation will reduce the real loan payments, and so the risk premium part of the interest rate can be correspondingly lower.

A number of ongoing or long-term business contracts also have provisions that prices will adjust automatically according to inflation. Sellers like such contracts because they are not locked into a low nominal selling price if inflation turns out higher than expected. Buyers like such contracts because they are not locked into a high buying price if inflation turns out to be lower than expected. A contract with automatic adjustments for inflation in effect agrees on a real price for the borrower to pay, rather than a nominal price.

Indexing in Government Programs

Many government programs are indexed to inflation. The U.S. income tax code is designed so that as a person's income rises above certain levels, the tax rate on the marginal income earned rises as well. That is what the expression “move into a higher tax bracket” means. For example, according to the basic tax tables from the Internal Revenue Service, in 2020, a married person filing jointly owed 10% of all taxable income from \$0 to \$19,750; 12% of all income from \$19,751 to \$80,250; 22% of all taxable income from \$80,251 to \$171,050; 24% of all taxable income from \$171,051 to \$326,600; 32% of all taxable income from \$326,601 to \$414,700; 35% of all taxable income from \$414,701 to \$622,050; and 37% of all income from \$622,051 and above.

Because of the many complex provisions in the rest of the tax code, it is difficult to determine exactly the taxes an individual owes the government based on these numbers, but the numbers illustrate the basic theme that tax rates rise as the marginal dollar of income rises. Until the late 1970s, if nominal wages increased along with inflation, people were moved into higher tax brackets and owed a higher proportion of their income in taxes, even though their real income had not risen. In 1981, the government eliminated this “bracket creep”. Now, the income levels where higher tax rates kick in are indexed to rise automatically with inflation.

The Social Security program offers two examples of indexing. Since the passage of the Social Security Indexing Act of 1972, the level of Social Security benefits increases each year along with the Consumer Price Index. Also, Social Security is funded by payroll taxes, which the government imposes on the income earned up to a certain amount—\$137,700 in 2020. The government adjusts this level of income upward each year according to the rate of inflation, so that an indexed increase in the Social Security tax base accompanies the indexed rise in the benefit level.

As yet another example of a government program affected by indexing, in 1996 the U.S. government began offering indexed bonds. Bonds are means by which the U.S. government (and many private-sector companies as well) borrows money; that is, investors buy the bonds, and then the government repays the money with interest. Traditionally, government bonds have paid a fixed rate of interest. This policy gave a government that had borrowed an incentive to encourage inflation, because it could then repay its past borrowing in inflated dollars at a lower real interest rate. However, indexed bonds promise to pay a certain real rate of interest above whatever inflation rate occurs. In the case of a retiree trying to plan for the long term and worried about the risk of inflation, for example, indexed bonds that guarantee a rate of return higher than inflation—no matter the level of inflation—can be a very comforting investment.

Might Indexing Reduce Concern over Inflation?

Indexing may seem like an obviously useful step. After all, when individuals, firms, and government programs are indexed against inflation, then people can worry less about arbitrary redistributions and other effects of inflation.

However, some of the fiercest opponents of inflation express grave concern about indexing. They point out that indexing is always partial. Not every employer will provide COLAs for workers. Not all companies can assume that costs and revenues will rise in lockstep with the general rates of inflation. Not all interest rates for borrowers and savers will change to match inflation exactly. However, as partial inflation indexing spreads, the political opposition to inflation may diminish. After all, older people whose Social Security benefits are protected against inflation, or banks that have loaned their money with adjustable-rate loans, no longer have as much reason to care whether inflation heats up. In a world where some people are indexed against inflation and some are not, financially savvy businesses and investors may seek out ways to be protected against inflation, while the financially unsophisticated and small businesses may feel it the most.

A Preview of Policy Discussions of Inflation

This chapter has focused on how economists measure inflation, historical experience with inflation, how to adjust nominal variables into real ones, how inflation affects the economy, and how indexing works. We have

barely hinted at the causes of inflation, and we have not addressed government policies to deal with inflation. We will examine these issues in depth in other chapters. However, it is useful to offer a preview here.

We can sum up the cause of inflation in one phrase: Too many dollars chasing too few goods. The great surges of inflation early in the twentieth century came after wars, which are a time when government spending is very high, but consumers have little to buy, because production is going to the war effort. Governments also commonly impose price controls during wartime. After the war, the price controls end and pent-up buying power surges forth, driving up inflation. Otherwise, if too few dollars are chasing too many goods, then inflation will decline or even turn into deflation. Therefore, we typically associate slowdowns in economic activity, as in major recessions and the Great Depression, with a reduction in inflation or even outright deflation.

The policy implications are clear. If we are to avoid inflation, the amount of purchasing power in the economy must grow at roughly the same rate as the production of goods. Macroeconomic policies that the government can use to affect the amount of purchasing power—through taxes, spending, and regulation of interest rates and credit—can thus cause inflation to rise or reduce inflation to lower levels.



BRING IT HOME

Inflation in a Pandemic—A Return to the 1970s, or a Temporary Adjustment?

The pandemic-induced recession caused all sorts of disruptions to our economy, including inflation. During the pandemic, the prices for goods like gas and cars fell as people shifted to remote work and canceled travel plans. But as the economy started to recover from the pandemic in early 2021, we started to see large increases in these prices. Higher prices were also fueled by the injections of cash into the economy through stimulus checks and unemployment benefits. The pandemic also caused shortages throughout the global supply chain, further pushing prices up (you'll learn more about this idea in a few chapters when we talk about aggregate supply and demand).

With headline annualized inflation rates in 2021 exceeding 6%, the question in the next few years is whether we'll see permanently high inflation rates of 9%, 10%, or more, like we did in the 1970s and early-1980s. Some economists believe the pandemic-induced inflation is just a transitory adjustment—indeed, used car and gasoline prices rose dramatically in 2010 and 2011 as well, as we were recovering from the Great Recession. Others are more concerned that the inflation is permanent. Shortages continue to exist throughout the economy as of early 2022, and if consumers expect higher inflation, it can be a self-fulfilling prophecy, as they start buying things now in order to avoid future bouts of inflation.

As you learned about earlier, inflation is a major concern of consumers, if less of an issue among economists. But inflation should be matched by increases in living standards, otherwise there could be major implications for the economy.

Key Terms

adjustable-rate mortgage (ARM) a loan a borrower uses to purchase a home in which the interest rate varies with market interest rates

base year arbitrary year whose value as an index number economists define as 100; inflation from the base year to other years can easily be seen by comparing the index number in the other year to the index number in the base year—for example, 100; so, if the index number for a year is 105, then there has been exactly 5% inflation between that year and the base year

basket of goods and services a hypothetical group of different items, with specified quantities of each one meant to represent a “typical” set of consumer purchases, used as a basis for calculating how the price level changes over time

Consumer Price Index (CPI) a measure of inflation that U.S. government statisticians calculate based on the price level from a fixed basket of goods and services that represents the average consumer's purchases

core inflation index a measure of inflation typically calculated by taking the CPI and excluding volatile economic variables such as food and energy prices to better measure the underlying and persistent trend in long-term prices

cost-of-living adjustments (COLAs) a contractual provision that wage increases will keep up with inflation

deflation negative inflation; most prices in the economy are falling

Employment Cost Index a measure of inflation based on wages paid in the labor market

GDP deflator a measure of inflation based on the prices of all the GDP components

hyperinflation an outburst of high inflation that often occurs (although not exclusively) when economies shift from a controlled economy to a market-oriented economy

index number a unit-free number derived from the price level over a number of years, which makes computing inflation rates easier, since the index number has values around 100

indexed a price, wage, or interest rate is adjusted automatically for inflation

inflation a general and ongoing rise in price levels in an economy

International Price Index a measure of inflation based on the prices of merchandise that is exported or imported

Producer Price Index (PPI) a measure of inflation based on prices paid for supplies and inputs by producers of goods and services

quality/new goods bias inflation calculated using a fixed basket of goods over time tends to overstate the true rise in cost of living, because it does not account for improvements in the quality of existing goods or the invention of new goods

substitution bias an inflation rate calculated using a fixed basket of goods over time tends to overstate the true rise in the cost of living, because it does not take into account that the person can substitute away from goods whose prices rise considerably

Key Concepts and Summary

22.1 Tracking Inflation

Economists measure the price level by using a basket of goods and services and calculating how the total cost of buying that basket of goods will increase over time. Economists often express the price level in terms of index numbers, which transform the cost of buying the basket of goods and services into a series of numbers in the same proportion to each other, but with an arbitrary base year of 100. We measure the inflation rate as the percentage change between price levels or index numbers over time.

22.2 How to Measure Changes in the Cost of Living

Measuring price levels with a fixed basket of goods will always have two problems: the substitution bias, by which a fixed basket of goods does not allow for buying more of what becomes relatively less expensive and less of what becomes relatively more expensive; and the quality/new goods bias, by which a fixed basket cannot account for improvements in quality and the advent of new goods. These problems can be reduced in

degree—for example, by allowing the basket of goods to evolve over time—but we cannot totally eliminate them. The most commonly cited measure of inflation is the Consumer Price Index (CPI), which is based on a basket of goods representing what the typical consumer buys. The Core Inflation Index further breaks down the CPI by excluding volatile economic commodities. Several price indices are not based on baskets of consumer goods. The GDP deflator is based on all GDP components. The Producer Price Index is based on prices of supplies and inputs bought by producers of goods and services. An Employment Cost Index measures wage inflation in the labor market. An International Price Index is based on the prices of merchandise that is exported or imported.

22.3 How the U.S. and Other Countries Experience Inflation

In the U.S. economy, the annual inflation rate in the last two decades has typically been around 2% to 4%. The periods of highest inflation in the United States in the twentieth century occurred during the years after World Wars I and II, and in the 1970s. The period of lowest inflation—actually, with deflation—was the 1930s Great Depression.

22.4 The Confusion Over Inflation

Unexpected inflation will tend to hurt those whose money received, in terms of wages and interest payments, does not rise with inflation. In contrast, inflation can help those who owe money that they can pay in less valuable, inflated dollars. Low rates of inflation have relatively little economic impact over the short term. Over the medium and the long term, even low rates of inflation can complicate future planning. High rates of inflation can muddle price signals in the short term and prevent market forces from operating efficiently, and can vastly complicate long-term savings and investment decisions.

22.5 Indexing and Its Limitations

A payment is indexed if it is automatically adjusted for inflation. Examples of indexing in the private sector include wage contracts with cost-of-living adjustments (COLAs) and loan agreements like adjustable-rate mortgages (ARMs). Examples of indexing in the public sector include tax brackets and Social Security payments.

Self-Check Questions

1. [Table 22.4](#) shows the fruit prices that the typical college student purchased from 2001 to 2004. What is the amount spent each year on the “basket” of fruit with the quantities shown in column 2?

Items	Qty	(2001) Price	(2001) Amount Spent	(2002) Price	(2002) Amount Spent	(2003) Price	(2003) Amount Spent	(2004) Price	(2004) Amount Spent
Apples	10	\$0.50		\$0.75		\$0.85		\$0.88	
Bananas	12	\$0.20		\$0.25		\$0.25		\$0.29	
Grapes	2	\$0.65		\$0.70		\$0.90		\$0.95	
Raspberries	1	\$2.00		\$1.90		\$2.05		\$2.13	\$2.13
Total									

TABLE 22.4

2. Construct the price index for a “fruit basket” in each year using 2003 as the base year.
3. Compute the inflation rate for fruit prices from 2001 to 2004.

4. Edna is living in a retirement home where most of her needs are taken care of, but she has some discretionary spending. Based on the basket of goods in Table 22.5, by what percentage does Edna's cost of living increase between time 1 and time 2?

Items	Quantity	(Time 1) Price	(Time 2) Price
Gifts for grandchildren	12	\$50	\$60
Pizza delivery	24	\$15	\$16
Blouses	6	\$60	\$50
Vacation trips	2	\$400	\$420

TABLE 22.5

5. [How to Measure Changes in the Cost of Living](#) introduced a number of different price indices. Which price index would be best to use to adjust your paycheck for inflation?
6. The Consumer Price Index is subject to the substitution bias and the quality/new goods bias. Are the Producer Price Index and the GDP Deflator also subject to these biases? Why or why not?
7. Go to this [website \(http://www.measuringworth.com/ppowerus/\)](http://www.measuringworth.com/ppowerus/) for the Purchasing Power Calculator at MeasuringWorth.com. How much money would it take today to purchase what one dollar would have bought in the year of your birth?
8. If inflation rises unexpectedly by 5%, would a state government that had recently borrowed money to pay for a new highway benefit or lose?
9. How should an increase in inflation affect the interest rate on an adjustable-rate mortgage?
10. A fixed-rate mortgage has the same interest rate over the life of the loan, whether the mortgage is for 15 or 30 years. By contrast, an adjustable-rate mortgage changes with market interest rates over the life of the mortgage. If inflation falls unexpectedly by 3%, what would likely happen to a homeowner with an adjustable-rate mortgage?

Review Questions

11. How do economists use a basket of goods and services to measure the price level?
12. Why do economists use index numbers to measure the price level rather than dollar value of goods?
13. What is the difference between the price level and the rate of inflation?
14. Why does “substitution bias” arise if we calculate the inflation rate based on a fixed basket of goods?
15. Why does the “quality/new goods bias” arise if we calculate the inflation rate based on a fixed basket of goods?
16. What has been a typical range of inflation in the U.S. economy in the last decade or so?
17. Over the last century, during what periods was the U.S. inflation rate highest and lowest?
18. What is deflation?
19. Identify several parties likely to be helped and hurt by inflation.
20. What is indexing?

21. Name several forms of indexing in the private and public sector.

Critical Thinking Questions

22. Inflation rates, like most statistics, are imperfect measures. Can you identify some ways that the inflation rate for fruit does not perfectly capture the rising price of fruit?
23. Given the federal budget deficit in recent years, some economists have argued that by adjusting Social Security payments for inflation using the CPI, Social Security is overpaying recipients. What is their argument, and do you agree or disagree with it?
24. Why is the GDP deflator not an accurate measure of inflation as it impacts a household?
25. Imagine that the government statisticians who calculate the inflation rate have been updating the basic basket of goods once every 10 years, but now they decide to update it every five years. How will this change affect the amount of substitution bias and quality/new goods bias?
26. Describe a situation, either a government policy situation, an economic problem, or a private sector situation, where using the CPI to convert from nominal to real would be more appropriate than using the GDP deflator.
27. Describe a situation, either a government policy situation, an economic problem, or a private sector situation, where using the GDP deflator to convert from nominal to real would be more appropriate than using the CPI.
28. Why do you think the U.S. experience with inflation over the last 50 years has been so much milder than in many other countries?
29. If, over time, wages and salaries on average rise at least as fast as inflation, why do people worry about how inflation affects incomes?
30. Who in an economy is the big winner from inflation?
31. If a government gains from unexpected inflation when it borrows, why would it choose to offer indexed bonds?
32. Do you think perfect indexing is possible? Why or why not?

Problems

33. The index number representing the price level changes from 110 to 115 in one year, and then from 115 to 120 the next year. Since the index number increases by five each year, is five the inflation rate each year? Is the inflation rate the same each year? Explain your answer.
34. The total price of purchasing a basket of goods in the United Kingdom over four years is: year 1=£940, year 2=£970, year 3=£1000, and year 4=£1070. Calculate two price indices, one using year 1 as the base year (set equal to 100) and the other using year 4 as the base year (set equal to 100). Then, calculate the inflation rate based on the first price index. If you had used the other price index, would you get a different inflation rate? If you are unsure, do the calculation and find out.
35. Within 1 or 2 percentage points, what has the U.S. inflation rate been during the last 20 years? Draw a graph to show the data.

- 36.** If inflation rises unexpectedly by 5%, indicate for each of the following whether the economic actor is helped, hurt, or unaffected:
- a. A union member with a COLA wage contract
 - b. Someone with a large stash of cash in a safe deposit box
 - c. A bank lending money at a fixed rate of interest
 - d. A person who is not due to receive a pay raise for another 11 months
- 37.** Rosalie the Retiree knows that when she retires in 16 years, her company will give her a one-time payment of \$20,000. However, if the inflation rate is 6% per year, how much buying power will that \$20,000 have when measured in today's dollars? *Hint:* Start by calculating the rise in the price level over the 16 years.

The International Trade and Capital Flows

23



FIGURE 23.1 A World of Money We are all part of the global financial system, which includes many different currencies. (Credit: modification of "Money from around the world" by Images Money/Flickr, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Measuring Trade Balances
- Trade Balances in Historical and International Context
- Trade Balances and Flows of Financial Capital
- The National Saving and Investment Identity
- The Pros and Cons of Trade Deficits and Surpluses
- The Difference between Level of Trade and the Trade Balance

Introduction to the International Trade and Capital Flows



BRING IT HOME

More than Meets the Eye in the Congo

How much do you interact with the global financial system? Do you think not much? Think again. Suppose you take

out a student loan, or you deposit money into your bank account. You just affected domestic savings and borrowing. Now say you are at the mall and buy two T-shirts “made in China,” and later contribute to a charity that helps refugees. What is the impact? You affected how much money flows into and out of the United States. If you open an IRA savings account and put money in an international mutual fund, you are involved in the flow of money overseas. While your involvement may not seem as influential as that of someone like the president, who can increase or decrease foreign aid and, thereby, have a huge impact on money flows in and out of the country, you do interact with the global financial system on a daily basis.

The balance of payments—a term you will meet soon—seems like a huge topic, but once you learn the specific components of trade and money, it all makes sense. Along the way, you may have to give up some common misunderstandings about trade and answer some questions: If a country is running a trade deficit, is that bad? Is a trade surplus good? For example, look at the Democratic Republic of the Congo (often referred to as “Congo”), a large country in Central Africa. In 2013, it ran a trade surplus of \$1 billion, so it must be doing well, right? In contrast, the trade deficit in the United States was \$508 billion in 2013. Do these figures suggest that the United States economy is performing worse than the Congolese economy? Not necessarily. The U.S. trade deficit tends to worsen as the economy strengthens. In contrast, high poverty rates in the Congo persist, and these rates are not going down even with the positive trade balance. Clearly, it is more complicated than simply asserting that running a trade deficit is bad for the economy. You will learn more about these issues and others in this chapter.

The **balance of trade** (or trade balance) is any gap between a nation’s dollar value of its exports, or what its producers sell abroad, and a nation’s dollar value of imports, or the foreign-made products and services that households and businesses purchase. Recall from [The Macroeconomic Perspective](#) that if exports exceed imports, the economy has a trade surplus. If imports exceed exports, the economy has a trade deficit. If exports and imports are equal, then trade is balanced, but what happens when trade is out of balance and large trade surpluses or deficits exist?

Germany, for example, has had substantial trade surpluses in recent decades, in which exports have greatly exceeded imports. According to the World Bank, in 2020, Germany ran a trade surplus of \$242 billion. In contrast, the U.S. economy in recent decades has experienced large trade deficits, in which imports have considerably exceeded exports. In 2020, for example, U.S. imports exceeded exports by \$651 billion.

A series of financial crises triggered by unbalanced trade can lead economies into deep recessions. These crises begin with large trade deficits. At some point, foreign investors become pessimistic about the economy and move their money to other countries. The economy then drops into deep recession, with real GDP often falling up to 10% or more in a single year. This happened to Mexico in 1995 when their GDP fell 8.1%. A number of countries in East Asia—Thailand, South Korea, Malaysia, and Indonesia—succumbed to the same economic illness in 1997–1998 (called the Asian Financial Crisis). In the late 1990s and into the early 2000s, Russia and Argentina had the identical experience. What are the connections between imbalances of trade in goods and services and the flows of international financial capital that set off these economic avalanches?

We will start by examining the balance of trade in more detail, by looking at some patterns of trade balances in the United States and around the world. Then we will examine the intimate connection between international flows of goods and services and international flows of financial capital, which to economists are really just two sides of the same coin. People often assume that trade surpluses like those in Germany must be a positive sign for an economy, while trade deficits like those in the United States must be harmful. As it turns out, both trade surpluses and deficits can be either good or bad. We will see why in this chapter.

23.1 Measuring Trade Balances

LEARNING OBJECTIVES

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

- Explain merchandise trade balance, current account balance, and unilateral transfers
- Identify components of the U.S. current account balance
- Calculate the merchandise trade balance and current account balance using import and export data for a country

A few decades ago, it was common to track the solid or physical items that planes, trains, and trucks transported between countries as a way of measuring the balance of trade. Economists call this measurement the **merchandise trade balance**. In most high-income economies, including the United States, goods comprise less than half of a country's total production, while services comprise more than half. The last two decades have seen a surge in international trade in services, powered by technological advances in telecommunications and computers that have made it possible to export or import customer services, finance, law, advertising, management consulting, software, construction engineering, and product design. Most global trade still takes the form of goods rather than services, and the government announces and the media prominently report the merchandise trade balance. Old habits are hard to break. Economists, however, typically rely on broader measures such as the balance of trade or the **current account balance** which includes other international flows of income and foreign aid.

Components of the U.S. Current Account Balance

[Table 23.1](#) breaks down the four main components of the U.S. current account balance for the last quarter of 2015 (seasonally adjusted). The first line shows the merchandise trade balance; that is, exports and imports of goods. Because imports exceed exports, the trade balance in the final column is negative, showing a merchandise trade deficit. We can explain how the government collects this trade information in the following Clear It Up feature.

	Value of Exports (money flowing into the United States)	Value of Imports (money flowing out of the United States)	Balance
Goods	\$1,428.8	\$2,350.8	−\$922.0
Services	\$705.6	\$460.3	\$245.3
Income receipts and payments	\$957.9	\$769.4	\$188.5
Unilateral transfers	\$166.3	\$294.2	−\$127.9
Current account balance	\$3,258.6	\$3,874.7	−\$616.1

TABLE 23.1 Components of the U.S. Current Account Balance for 2020 (in billions of dollars). Source: <https://apps.bea.gov/itable/index.cfm>, ITA Table 1.1.



CLEAR IT UP

How does the U.S. government collect trade statistics?

Do not confuse the balance of trade (which tracks imports and exports), with the current account balance, which includes not just exports and imports, but also income from investment and transfers.

The Bureau of Economic Analysis (BEA) within the U.S. Department of Commerce compiles statistics on the balance of trade using a variety of different sources. Merchandise importers and exporters must file monthly documents with the Census Bureau, which provides the basic data for tracking trade. To measure international trade in services—which can happen over a telephone line or computer network without shipping any physical goods—the BEA carries out a set of surveys. Another set of BEA surveys tracks investment flows, and there are even specific surveys to collect travel information from U.S. residents visiting Canada and Mexico. For measuring unilateral transfers, the BEA has access to official U.S. government spending on aid, and then also carries out a survey of charitable organizations that make foreign donations.

The BEA then cross-checks this information on international flows of goods and capital against other available data. For example, the Census Bureau also collects data from the shipping industry, which it can use to check the data on trade in goods. All companies involved in international flows of capital—including banks and companies making financial investments like stocks—must file reports, which the U.S. Department of the Treasury ultimately checks. The BEA also can cross check information on foreign trade by looking at data collected by other countries on their foreign trade with the United States, and also at the data collected by various international organizations. Take these data sources, stir carefully, and you have the U.S. balance of trade statistics. Much of the statistics that we cite in this chapter come from these sources.

The second row of [Table 23.1](#) provides data on trade in services. Here, the U.S. economy is running a surplus. Although the level of trade in services is still relatively small compared to trade in goods, the importance of services has expanded substantially over the last few decades. For example, U.S. exports of services were equal to about one-half of U.S. exports of goods in 2020, compared to one-fifth in 1980.

The third component of the current account balance, labeled “income payments,” refers to money that U.S. financial investors received on their foreign investments (money flowing into the United States) and payments to foreign investors who had invested their funds here (money flowing out of the United States). The reason for including this money on foreign investment in the overall measure of trade, along with goods and services, is that, from an economic perspective, income is just as much an economic transaction as car, wheat, or oil shipments: it is just trade that is happening in the financial capital market.

The final category of the current account balance is **unilateral transfers**, which are payments that government, private charities, or individuals make in which they send money abroad without receiving any direct good or service. Economic or military assistance from the U.S. government to other countries fits into this category, as does spending abroad by charities to address poverty or social inequalities. When an individual in the United States sends money overseas, as is the case with some immigrants, it is also counted in this category. The current account balance treats these unilateral payments like imports, because they also involve a stream of payments leaving the country. For the U.S. economy, unilateral transfers are almost always negative. This pattern, however, does not always hold. In 1991, for example, when the United States led an international coalition against Saddam Hussein’s Iraq in the Gulf War, many other nations agreed that they would make payments to the United States to offset the U.S. war expenses. These payments were large enough that, in 1991, the overall U.S. balance on unilateral transfers was a positive \$10 billion.

The following Work It Out feature steps you through the process of using the values for goods, services, and income payments to calculate the merchandise balance and the current account balance.

WORK IT OUT

Calculating the Merchandise Balance and the Current Account Balance

	Exports (in \$ billions)	Imports (in \$ billions)	Balance
Goods			
Services			
Income payments			
Unilateral transfers			
Current account balance			

TABLE 23.2 Calculating Merchandise Balance and Current Account Balance

Use the information given below to fill in [Table 23.2](#), and then calculate:

- The merchandise balance
- The current account balance

Known information:

- Unilateral transfers: \$130
- Exports in goods: \$1,046
- Exports in services: \$509
- Imports in goods: \$1,562
- Imports in services: \$371
- Income received by U.S. investors on foreign stocks and bonds: \$561
- Income received by foreign investors on U.S. assets: \$472

Step 1. Focus on goods and services first. Enter the dollar amount of exports of both goods and services under the Export column.

Step 2. Enter imports of goods and services under the Import column.

Step 3. Under the Export column and in the row for Income payments, enter the financial flows of money coming back to the United States. U.S. investors are earning this income from abroad.

Step 4. Under the Import column and in the row for Income payments, enter the financial flows of money going out of the United States to foreign investors. Foreign investors are earning this money on U.S. assets, like stocks.

Step 5. Unilateral transfers are money flowing out of the United States in the form of, for example, military aid, foreign aid, and global charities. Because the money leaves the country, enter it under Imports and in the final column as well, as a negative.

Step 6. Calculate the trade balance by subtracting imports from exports in both goods and services. Enter this in the final Balance column. This can be positive or negative.

Step 7. Subtract the income payments flowing out of the country (under Imports) from the money coming back to the United States (under Exports) and enter this amount under the Balance column.

Step 8. Enter unilateral transfers as a negative amount under the Balance column.

Step 9. The merchandise trade balance is the difference between exports of goods and imports of goods—the

first number under Balance.

Step 10. Now sum up your columns for Exports, Imports, and Balance. The final balance number is the current account balance.

The merchandise balance of trade is the difference between exports and imports. In this case, it is equal to $\$1,046 - \$1,562$, a trade deficit of $-\$516$ billion. The current account balance is $-\$419$ billion. See the completed [Table 23.3](#).

	Value of Exports (money flowing into the United States)	Value of Imports (money flowing out of the United States)	Balance
Goods	\$1,046	\$1,562	-\$516
Services	\$509	\$371	\$138
Income receipts and payments	\$561	\$472	\$89
Unilateral transfers	\$0	\$130	-\$130
Current account balance	\$2,116	\$2,535	-\$419

TABLE 23.3 Completed Merchandise Balance and Current Account Balance

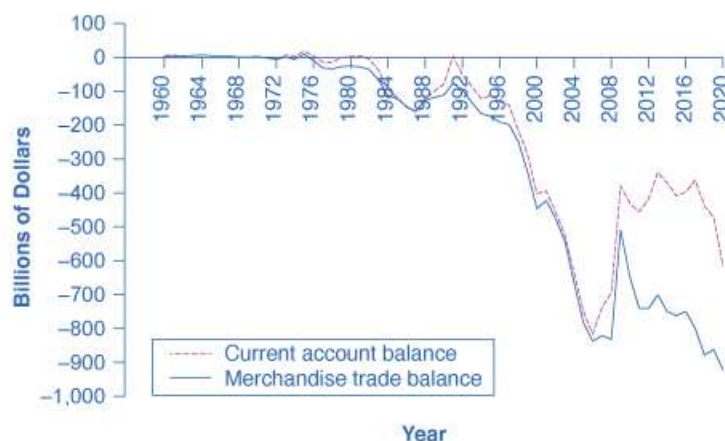
23.2 Trade Balances in Historical and International Context

LEARNING OBJECTIVES

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

- Analyze graphs of the current account balance and the merchandise trade balance
- Identify patterns in U.S. trade surpluses and deficits
- Compare the U.S. trade surpluses and deficits to other countries' trade surpluses and deficits

We present the history of the U.S. current account balance in recent decades in several different ways. [Figure 23.2](#) (a) shows the current account balance and the merchandise trade balance—the latter of which is simply the balance on goods exported versus imported—in dollar terms. [Figure 23.2](#) (b) shows the current account balance and merchandise account balance yet again, this time as a share of the GDP for that year. By dividing the trade deficit in each year by GDP in that year, [Figure 23.2](#) (b) factors out both inflation and growth in the real economy.



(a) The current account and merchandise trade balance in nominal dollars



(b) The current account and merchandise trade balance as a percentage of GDP

FIGURE 23.2 Current Account Balance and Merchandise Trade Balance, 1960–2020 (a) The current account balance and the merchandise trade balance in billions of dollars from 1960 to 2020. The merchandise trade balance is the trade balance on goods only. If the lines are above zero dollars, the United States was running a positive trade balance and current account balance. If the lines fall below zero dollars, the United States is running a trade deficit and a deficit in its current account balance. (b) This shows the same items—trade balance and current account balance—in relationship to the size of the U.S. economy, or GDP, from 1960 to 2020.

By either measure, the U.S. balance of trade pattern is clear. From the 1960s into the 1970s, the U.S. economy had mostly small trade surpluses—that is, the graphs in [Figure 23.2](#) show positive numbers. However, starting in the 1980s, the trade deficit increased rapidly, and after a tiny surplus in 1991, the current account trade deficit became even larger in the late 1990s and into the mid-2000s. However, the trade deficit declined in 2009 after the recession had taken hold, then rebounded partially in 2010 and remained stable up through 2019, before falling again in 2020.

[Click to view content \(https://openstax.org/books/principles-economics-3e/pages/23-2-trade-balances-in-historical-and-international-context\)](https://openstax.org/books/principles-economics-3e/pages/23-2-trade-balances-in-historical-and-international-context)

Current Account Balance in Billions of Dollars.

Table 23.4 shows the U.S. trade picture in 2013 compared with some other economies from around the world. While the U.S. economy has consistently run trade deficits in recent years, Japan and many European nations, among them France and Germany, have consistently run trade surpluses. Some of the other countries listed include Brazil, the largest economy in Latin America; Nigeria, along with South Africa competing to be the largest economy in Africa; and China, India, and Korea. The first column offers one measure of an economy's globalization: **exports of goods and services as a percentage of GDP**. The second column shows the trade balance. Usually, most countries have trade surpluses or deficits that are less than 5% of GDP. As you can see, the U.S. current account balance is -2.6% of GDP, while Germany's is 8.4% of GDP.

	Exports of Goods and Services	Current Account Balance
United States	10.2%	-2.9%
Japan	15.5%	3.2%
Germany	43.4%	7.0%
United Kingdom	27.9%	-2.6%
Canada	29.0%	-1.8%
Sweden	44.6%	5.7%
Korea	36.4%	4.6%
Mexico	40.2%	2.4%
Brazil	16.9%	-1.8%
China	18.5%	1.9%
India	18.7%	1.2%
Nigeria	8.8%	-3.9%
World	-	0.0%

TABLE 23.4 Level and Balance of Trade (Balance of Payments basis) in 2020
(figures as a percentage of GDP, Source: <http://data.worldbank.org/indicator/BN.CAB.XOKA.GD.ZS>)

23.3 Trade Balances and Flows of Financial Capital

LEARNING OBJECTIVES

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

- Explain the connection between trade balances and financial capital flows
- Calculate comparative advantage
- Explain balanced trade in terms of investment and capital flows

As economists see it, trade surpluses can be either good or bad, depending on circumstances, and trade deficits can be good or bad, too. The challenge is to understand how the international flows of goods and services are connected with international flows of **financial capital**. In this module we will illustrate the intimate connection between trade balances and flows of financial capital in two ways: a parable of trade between Robinson Crusoe and Friday, and a circular flow diagram representing flows of trade and payments.

A Two-Person Economy: Robinson Crusoe and Friday

To understand how economists view trade deficits and surpluses, consider a parable based on the story of Robinson Crusoe. Crusoe, as you may remember from the classic novel by Daniel Defoe first published in 1719, was shipwrecked on a desert island. After living alone for some time, he is joined by a second person, whom he names Friday. Think about the balance of trade in a two-person economy like that of Robinson and Friday.

Robinson and Friday trade goods and services. Perhaps Robinson catches fish and trades them to Friday for coconuts, or Friday weaves a hat out of tree fronds and trades it to Robinson for help in carrying water. For a period of time, each individual trade is self-contained and complete. Because each trade is voluntary, both Robinson and Friday must feel that they are receiving fair value for what they are giving. As a result, each person's exports are always equal to his imports, and trade is always in balance between the two. Neither person experiences either a trade deficit or a trade surplus.

However, one day Robinson approaches Friday with a proposition. Robinson wants to dig ditches for an irrigation system for his garden, but he knows that if he starts this project, he will not have much time left to fish and gather coconuts to feed himself each day. He proposes that Friday supply him with a certain number of fish and coconuts for several months, and then after that time, he promises to repay Friday out of the extra produce that he will be able to grow in his irrigated garden. If Friday accepts this offer, then a trade imbalance comes into being. For several months, Friday will have a trade surplus: that is, he is exporting to Robinson more than he is importing. More precisely, he is giving Robinson fish and coconuts, and at least for the moment, he is receiving nothing in return. Conversely, Robinson will have a trade deficit, because he is importing more from Friday than he is exporting.

This parable raises several useful issues in thinking about what a trade deficit and a trade surplus really mean in economic terms. The first issue that this story of Robinson and Friday raises is this: Is it better to have a trade surplus or a trade deficit? The answer, as in any voluntary market interaction, is that if both parties agree to the transaction, then they may both be better off. Over time, if Robinson's irrigated garden is a success, it is certainly possible that both Robinson and Friday can benefit from this agreement.

The parable raises a second issue: What can go wrong? Robinson's proposal to Friday introduces an element of uncertainty. Friday is, in effect, making a loan of fish and coconuts to Robinson, and Friday's happiness with this arrangement will depend on whether Robinson repays that loan as planned, in full and on time. Perhaps Robinson spends several months loafing and never builds the irrigation system, or perhaps Robinson has been too optimistic about how much he will be able to grow with the new irrigation system, which turns out not to be very productive. Perhaps, after building the irrigation system, Robinson decides that he does not want to repay Friday as much as he previously agreed. Any of these developments will prompt a new round of negotiations between Friday and Robinson. Why the repayment failed is likely to shape Friday's attitude toward these renegotiations. If Robinson worked very hard and the irrigation system just did not increase production as intended, Friday may have some sympathy. If Robinson loafed or if he just refuses to pay, Friday may become irritated.

A third issue that the parable raises is that an intimate relationship exists between a trade deficit and international borrowing, and between a trade surplus and international lending. The size of Friday's trade surplus is exactly how much he is lending to Robinson. The size of Robinson's trade deficit is exactly how much he is borrowing from Friday. To economists, a trade surplus literally means the same thing as an outflow of financial capital, and a trade deficit literally means the same thing as an inflow of financial capital. This last insight is worth exploring in greater detail, which we will do in the following section.

The story of Robinson and Friday also provides a good opportunity to consider the law of comparative advantage, which you learn more about in the [International Trade](#) chapter. The following Work It Out feature steps you through calculating comparative advantage for the wheat and cloth traded between the United States and Great Britain in the 1800s.

WORK IT OUT

Calculating Comparative Advantage

In the 1800s, the United States and Britain traded wheat and cloth. [Table 23.5](#) shows the varying hours of labor per unit of output.

	Wheat (in bushels)	Cloth (in yards)	Relative labor cost of wheat (Pw/Pc)	Relative labor cost of cloth (Pc/Pw)
United States	8	9	8/9	9/8
Britain	4	3	4/3	3/4

TABLE 23.5

Step 1. Observe from [Table 23.5](#) that, in the United States, it takes eight hours to supply a bushel of wheat and nine hours to supply a yard of cloth. In contrast, it takes four hours to supply a bushel of wheat and three hours to supply a yard of cloth in Britain.

Step 2. Recognize the difference between absolute advantage and comparative advantage. Britain has an absolute advantage (lowest cost) in each good, since it takes a lower amount of labor to make each good in Britain. Britain also has a comparative advantage in the production of cloth (lower opportunity cost in cloth (3/4 versus 9/8)). The United States has a comparative advantage in wheat production (lower opportunity cost of 8/9 versus 4/3).

Step 3. Determine the relative price of one good in terms of the other good. The price of wheat, in this example, is the amount of cloth you have to give up. To find this price, convert the hours per unit of wheat and cloth into units per hour. To do so, observe that in the United States it takes eight hours to make a bushel of wheat, so workers can process 1/8 of a bushel of wheat in an hour. It takes nine hours to make a yard of cloth in the United States, so workers can produce 1/9 of a yard of cloth in an hour. If you divide the amount of cloth (1/9 of a yard) by the amount of wheat you give up (1/8 of a bushel) in an hour, you find the price (8/9) of one good (wheat) in terms of the other (cloth).

The Balance of Trade as the Balance of Payments

The connection between trade balances and international flows of financial capital is so close that economists sometimes describe the balance of trade as the balance of payments. Each category of the current account balance involves a corresponding flow of payments between a given country and the rest of the world economy.

[Figure 23.3](#) shows the flow of goods and services and payments between one country—the United States in this example—and the rest of the world. The top line shows U.S. exports of goods and services, while the second line shows financial payments from purchasers in other countries back to the U.S. economy. The third line then shows U.S. imports of goods, services, and investment, and the fourth line shows payments from the home economy to the rest of the world. Flow of goods and services (lines one and three) show up in the current account, while we find flow of funds (lines two and four) in the financial account.

The bottom four lines in [Figure 23.3](#) show the flows of investment income. In the first of the bottom lines, we see investments made abroad with funds flowing from the home country to the rest of the world. Investment income stemming from an investment abroad then runs in the other direction from the rest of the world to the home country. Similarly, we see on the bottom third line, an investment from the rest of the world into the

home country and investment income (bottom fourth line) flowing from the home country to the rest of the world. We find the investment income (bottom lines two and four) in the current account, while investment to the rest of the world or into the home country (lines one and three) is in the financial account. This figure does not show unilateral transfers, the fourth item in the current account.

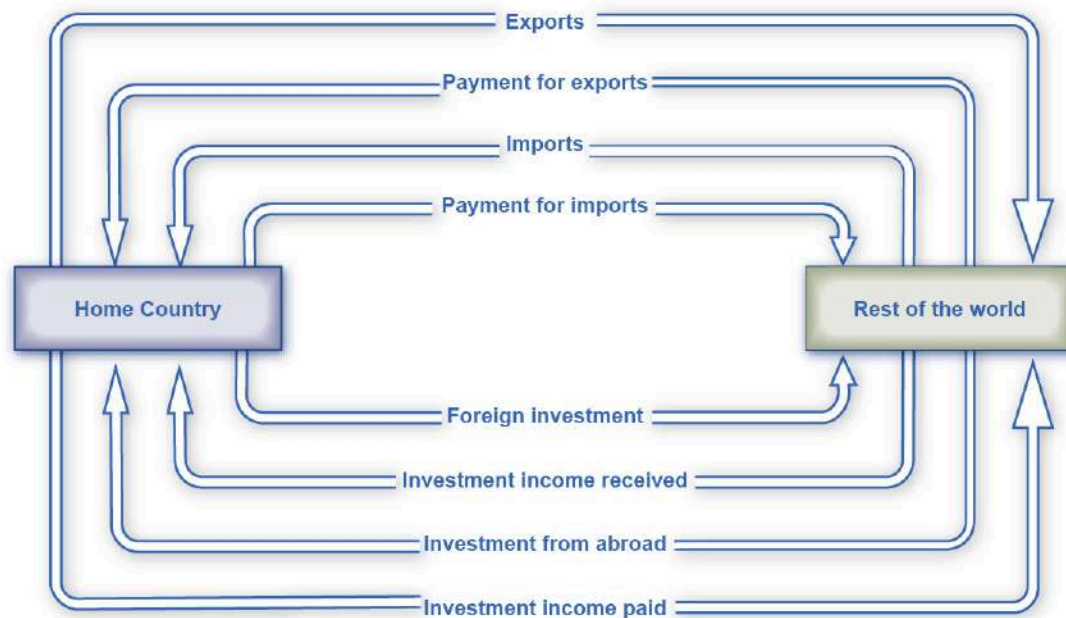


FIGURE 23.3 Flow of Investment Goods and Capital Each element of the current account balance involves a flow of financial payments between countries. The top line shows exports of goods and services leaving the home country; the second line shows the money that the home country receives for those exports. The third line shows imports that the home country receives; the fourth line shows the payments that the home country sent abroad in exchange for these imports.

A current account deficit means that, the country is a net borrower from abroad. Conversely, a positive current account balance means a country is a net lender to the rest of the world. Just like the parable of Robinson and Friday, the lesson is that a trade surplus means an overall outflow of financial investment capital, as domestic investors put their funds abroad, while a deficit in the current account balance is exactly equal to the overall or net inflow of foreign investment capital from abroad.

It is important to recognize that an inflow and outflow of foreign capital does not necessarily refer to a debt that governments owe to other governments, although government debt may be part of the picture. Instead, these international flows of financial capital refer to all of the ways in which private investors in one country may invest in another country—by buying real estate, companies, and financial investments like stocks and bonds.

23.4 The National Saving and Investment Identity

LEARNING OBJECTIVES

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

- Explain the determinants of trade and current account balance
- Identify and calculate supply and demand for financial capital
- Explain how a nation's own level of domestic saving and investment determines a nation's balance of trade
- Predict the rising and falling of trade deficits based on a nation's saving and investment identity

The close connection between trade balances and international flows of savings and investments leads to a macroeconomic analysis. This approach views trade balances—and their associated flows of financial

capital—in the context of the overall levels of savings and financial investment in the economy.

Understanding the Determinants of the Trade and Current Account Balance

The **national saving and investment identity** provides a useful way to understand the determinants of the trade and current account balance. In a nation's financial capital market, the quantity of financial capital supplied at any given time must equal the quantity of financial capital demanded for purposes of making investments. What is on the supply and demand sides of financial capital? See the following Clear It Up feature for the answer to this question.



CLEAR IT UP

What comprises the supply and demand of financial capital?

A country's national savings is the total of its domestic savings by household and companies (private savings) as well as the government (public savings). If a country is running a trade deficit, it means money from abroad is entering the country and the government considers it part of the supply of financial capital.

The demand for financial capital (money) represents groups that are borrowing the money. Businesses need to borrow to finance their investments in factories, materials, and personnel. When the federal government runs a budget deficit, it is also borrowing money from investors by selling Treasury bonds. Therefore, both business investment and the federal government can demand (or borrow) the supply of savings.

There are two main sources for the supply of financial capital in the U.S. economy: saving by individuals and firms, called S , and the inflow of financial capital from foreign investors, which is equal to the trade deficit ($M - X$), or imports minus exports. There are also two main sources of demand for financial capital in the U.S. economy: private sector investment, I , and government borrowing, where the government needs to borrow when government spending, G , is higher than the taxes collected, T . We can express this national savings and investment identity in algebraic terms:

$$\begin{aligned}\text{Supply of financial capital} &= \text{Demand for financial capital} \\ S + (M - X) &= I + (G - T)\end{aligned}$$

Again, in this equation, S is private savings, T is taxes, G is government spending, M is imports, X is exports, and I is investment. This relationship is true as a matter of definition because, for the macro economy, the quantity supplied of financial capital must be equal to the quantity demanded.

However, certain components of the national savings and investment identity can switch between the supply side and the demand side. Some countries, like the United States in most years since the 1970s, have budget deficits, which mean the government is spending more than it collects in taxes, and so the government needs to borrow funds. In this case, the government term would be $G - T > 0$, showing that spending is larger than taxes, and the government would be a demander of financial capital on the left-hand side of the equation (that is, a borrower), not a supplier of financial capital on the right-hand side. However, if the government runs a budget surplus so that the taxes exceed spending, as the U.S. government did from 1998 to 2001, then the government in that year was contributing to the supply of financial capital ($T - G > 0$), and would appear on the left (saving) side of the national savings and investment identity.

Similarly, if a national economy runs a trade surplus, the trade sector will involve an outflow of financial capital to other countries. A trade surplus means that the domestic financial capital is in surplus within a country and can be invested in other countries.

The fundamental notion that total quantity of financial capital demanded equals total quantity of financial capital supplied must always remain true. Domestic savings will always appear as part of the supply of financial capital and domestic investment will always appear as part of the demand for financial capital.

However, the government and trade balance elements of the equation can move back and forth as either suppliers or demanders of financial capital, depending on whether government budgets and the trade balance are in surplus or deficit.

Domestic Saving and Investment Determine the Trade Balance

One insight from the national saving and investment identity is that a nation's own levels of domestic saving and investment determine a nation's balance of trade. To understand this point, rearrange the identity to put the balance of trade all by itself on one side of the equation. Consider first the situation with a trade deficit, and then the situation with a trade surplus.

In the case of a trade deficit, the national saving and investment identity can be rewritten as:

$$\begin{aligned}\text{Trade deficit} &= \text{Domestic investment} - \text{Private domestic saving} - \text{Government (or public) savings} \\ (M - X) &= I - S - (T - G)\end{aligned}$$

In this case, domestic investment is higher than domestic saving, including both private and government saving. The only way that domestic investment can exceed domestic saving is if capital is flowing into a country from abroad. After all, that extra financial capital for investment has to come from someplace.

Now consider a trade *surplus* from the standpoint of the national saving and investment identity:

$$\begin{aligned}\text{Trade surplus} &= \text{Private domestic saving} + \text{Public saving} - \text{Domestic investment} \\ (X - M) &= S + (T - G) - I\end{aligned}$$

In this case, domestic savings (both private and public) is higher than domestic investment. That extra financial capital will be invested abroad.

This connection of domestic saving and investment to the trade balance explains why economists view the balance of trade as a fundamentally macroeconomic phenomenon. As the national saving and investment identity shows, the performance of certain sectors of an economy, like cars or steel, do not determine the trade balance. Further, whether the nation's trade laws and regulations encourage free trade or protectionism also does not determine the trade balance (see [Globalization and Protectionism](#)).

Exploring Trade Balances One Factor at a Time

The national saving and investment identity also provides a framework for thinking about what will cause trade deficits to rise or fall. Begin with the version of the identity that has domestic savings and investment on the left and the trade deficit on the right:

$$\begin{aligned}\text{Domestic investment} - \text{Private domestic savings} - \text{Public domestic savings} &= \text{Trade deficit} \\ I - S - (T - G) &= (M - X)\end{aligned}$$

Now, consider the factors on the left-hand side of the equation one at a time, while holding the other factors constant.

As a first example, assume that the level of domestic investment in a country rises, while the level of private and public saving remains unchanged. [Table 23.6](#) shows the result in the first row under the equation. Since the equality of the national savings and investment identity must continue to hold—it is, after all, an identity that must be true by definition—the rise in domestic investment will mean a higher trade deficit. This situation occurred in the U.S. economy in the late 1990s. Because of the surge of new information and communications technologies that became available, business investment increased substantially. A fall in private saving during this time and a rise in government saving more or less offset each other. As a result, the financial capital to fund that business investment came from abroad, which is one reason for the very high U.S. trade deficits of the late 1990s and early 2000s.

Domestic Investment	–	Private Domestic Savings	–	Public Domestic Savings	=	Trade Deficit
I	–	S	–	(T – G)	=	(M – X)
Up		No change		No change		Then M – X must rise
No change		Up		No change		Then M – X must fall
No change		No change		Down		Then M – X must rise

TABLE 23.6 Causes of a Changing Trade Balance

As a second scenario, assume that the level of domestic savings rises, while the level of domestic investment and public savings remain unchanged. In this case, the trade deficit would decline. As domestic savings rises, there would be less need for foreign financial capital to meet investment needs. For this reason, a policy proposal often made for reducing the U.S. trade deficit is to increase private saving—although exactly how to increase the overall rate of saving has proven controversial.

As a third scenario, imagine that the government budget deficit increased dramatically, while domestic investment and private savings remained unchanged. This scenario occurred in the U.S. economy in the mid-1980s. The federal budget deficit increased from \$79 billion in 1981 to \$221 billion in 1986—an increase in the demand for financial capital of \$142 billion. The current account balance collapsed from a surplus of \$5 billion in 1981 to a deficit of \$147 billion in 1986—an increase in the supply of financial capital from abroad of \$152 billion. The connection at that time is clear: a sharp increase in government borrowing increased the U.S. economy's demand for financial capital, and foreign investors through the trade deficit primarily supplied that increase. The following Work It Out feature walks you through a scenario in which private domestic savings has to rise by a certain amount to reduce a trade deficit.

WORK IT OUT**Solving Problems with the Saving and Investment Identity**

Use the saving and investment identity to answer the following question: Country A has a trade deficit of \$200 billion, private domestic savings of \$500 billion, a government deficit of \$200 billion, and private domestic investment of \$500 billion. To reduce the \$200 billion trade deficit by \$100 billion, by how much does private domestic savings have to increase?

Step 1. Write out the savings investment formula solving for the trade deficit or surplus on the left:

$$(X - M) = S + (T - G) - I$$

Step 2. In the formula, put the amount for the trade deficit in as a negative number ($X - M$). The left side of your formula is now:

$$-200 = S + (T - G) - I$$

Step 3. Enter the private domestic savings (S) of \$500 in the formula:

$$-200 = 500 + (T - G) - I$$

Step 4. Enter the private domestic investment (I) of \$500 into the formula:

$$-200 = 500 + (T - G) - 500$$

Step 5. The government budget surplus or balance is represented by $(T - G)$. Enter a budget deficit amount for $(T - G)$ of -200 :

$$-200 = 500 + (-200) - 500$$

Step 6. Your formula now is:

$$(X - M) = S + (T - G) - I$$

$$-200 = 500 + (-200) - 500$$

The question is: To reduce your trade deficit $(X - M)$ of -200 to -100 (in billions of dollars), by how much will savings have to rise?

$$(X - M) = S + (T - G) - I$$

$$-100 = S + (-200) - 500$$

$$600 = S$$

Step 7. Summarize the answer: Private domestic savings needs to rise by \$100 billion, to a total of \$600 billion, for the two sides of the equation to remain equal ($-100 = -100$).

Short-Term Movements in the Business Cycle and the Trade Balance

In the short run, whether an economy is in a recession or on the upswing can affect trade imbalances. A recession tends to make a trade deficit smaller, or a trade surplus larger, while a period of strong economic growth tends to make a trade deficit larger, or a trade surplus smaller. These are not hard-and-fast rules, however.

As an example, note in [Figure 23.2](#) that the U.S. trade deficit declined by almost half from 2006 to 2009. One primary reason for this change is that during the recession, as the U.S. economy slowed down, it purchased fewer of all goods, including fewer imports from abroad. However, buying power abroad fell less, and so U.S. exports did not fall by as much. On the other hand, during the 2020 pandemic-induced recession, the trade deficit expanded (current account balance plummeted) as the U.S. economy became more reliant on other countries for goods and services.

Conversely, in the mid-2000s, when the U.S. trade deficit became very large, a contributing short-term reason is that the U.S. economy was growing. As a result, there was considerable aggressive buying in the U.S. economy, including the buying of imports. Thus, a trade deficit (or a much lower trade surplus) often accompanies a rapidly growing domestic economy, while a trade surplus (or a much lower trade deficit) accompanies a slowing or recessionary domestic economy.

Regardless of whether the trade deficit falls or rises during expansions or downturns, when the trade deficit rises, it necessarily means a greater net inflow of foreign financial capital. The national saving and investment identity teaches that the rest of the economy can absorb this inflow of foreign financial capital in several different ways. For example, reduced private savings could offset the additional inflow of financial capital from abroad, leaving domestic investment and public saving unchanged. Alternatively, the inflow of foreign financial capital could result in higher domestic investment, leaving private and public saving unchanged. Yet another possibility is that greater government borrowing could absorb the inflow of foreign financial capital, leaving domestic saving and investment unchanged. The national saving and investment identity does not specify which of these scenarios, alone or in combination, will occur—only that one of them must occur.

23.5 The Pros and Cons of Trade Deficits and Surpluses

LEARNING OBJECTIVES

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

- Identify three ways in which borrowing money or running a trade deficit can result in a healthy economy
- Identify three ways in which borrowing money or running a trade deficit can result in a weaker economy

Because flows of trade always involve flows of financial payments, flows of international trade are actually the same as flows of international financial capital. The question of whether trade deficits or surpluses are good or bad for an economy is, in economic terms, exactly the same question as whether it is a good idea for an economy to rely on net inflows of financial capital from abroad or to make net investments of financial capital abroad. Conventional wisdom often holds that borrowing money is foolhardy, and that a prudent country, like a prudent person, should always rely on its own resources. While it is certainly possible to borrow too much—as anyone with an overloaded credit card can testify—borrowing at certain times can also make sound economic sense. For both individuals and countries, there is no economic merit in a policy of abstaining from participation in financial capital markets.

It makes economic sense to borrow when you are buying something with a long-run payoff; that is, when you are making an investment. For this reason, it can make economic sense to borrow for a college education, because the education will typically allow you to earn higher wages, and so to repay the loan and still come out ahead. It can also make sense for a business to borrow in order to purchase a machine that will last 10 years, as long as the machine will increase output and profits by more than enough to repay the loan. Similarly, it can make economic sense for a national economy to borrow from abroad, as long as it wisely invests the money in ways that will tend to raise the nation's economic growth over time. Then, it will be possible for the national economy to repay the borrowed money over time and still end up better off than before.

One vivid example of a country that borrowed heavily from abroad, invested wisely, and did perfectly well is the United States during the nineteenth century. The United States ran a trade deficit in 40 of the 45 years from 1831 to 1875, which meant that it was importing capital from abroad over that time. However, that financial capital was mostly invested in projects like railroads that brought a substantial economic payoff. (See the following Clear It Up feature for more on this.)

A more recent example along these lines is the experience of South Korea, which had trade deficits during much of the 1970s—and so was an importer of capital over that time. However, South Korea also had high rates of investment in physical plant and equipment, and its economy grew rapidly. From the mid-1980s into the mid-1990s, South Korea often had trade surpluses—that is, it was repaying its past borrowing by sending capital abroad.

In contrast, some countries have run large trade deficits, borrowed heavily in global capital markets, and ended up in all kinds of trouble. Two specific sorts of trouble are worth examining. First, a borrower nation can find itself in a bind if it does not invest the incoming funds from abroad in a way that leads to increased productivity. Several of Latin America's large economies, including Mexico and Brazil, ran large trade deficits and borrowed heavily from abroad in the 1970s, but the inflow of financial capital did not boost productivity sufficiently, which meant that these countries faced enormous troubles repaying the money borrowed when economic conditions shifted during the 1980s. Similarly, it appears that a number of African nations that borrowed foreign funds in the 1970s and 1980s did not invest in productive economic assets. As a result, several of those countries later faced large interest payments, with no economic growth to show for the borrowed funds.

**CLEAR IT UP****Are trade deficits always harmful?**

For most years of the nineteenth century, U.S. imports exceeded exports and the U.S. economy had a trade deficit. Yet the string of trade deficits did not hold back the economy at all. Instead, the trade deficits contributed to the strong economic growth that gave the U.S. economy the highest per capita GDP in the world by around 1900.

The U.S. trade deficits meant that the U.S. economy was receiving a net inflow of foreign capital from abroad. Much of that foreign capital flowed into two areas of investment—railroads and public infrastructure like roads, water systems, and schools—which were important to helping the U.S. economy grow.

We should not overstate the effect of foreign investment capital on U.S. economic growth. In most years the foreign financial capital represented no more than 6–10% of the funds that the government used for overall physical investment in the economy. Nonetheless, the trade deficit and the accompanying investment funds from abroad were clearly a help, not a hindrance, to the U.S. economy in the nineteenth century.

A second “trouble” is: What happens if the foreign money flows in, and then suddenly flows out again? We raised this scenario at the start of the chapter. In the mid-1990s, a number of countries in East Asia—Thailand, Indonesia, Malaysia, and South Korea—ran large trade deficits and imported capital from abroad. However, in 1997 and 1998 many foreign investors became concerned about the health of these economies, and quickly pulled their money out of stock and bond markets, real estate, and banks. The extremely rapid departure of that foreign capital staggered the banking systems and economies of these countries, plunging them into deep recession. We investigate and discuss the links between international capital flows, banks, and recession in [The Impacts of Government Borrowing](#).

While a trade deficit is not always harmful, there is no guarantee that running a trade surplus will bring robust economic health. For example, Germany and Japan ran substantial trade surpluses for most of the last three decades. Regardless of their persistent trade surpluses, both countries have experienced occasional recessions and neither country has had especially robust annual growth in recent years. Read more about Japan’s trade surplus in the next Clear It Up feature.

**LINK IT UP**

Watch this [video \(http://openstax.org/l/tradedeficit\)](http://openstax.org/l/tradedeficit) on whether or not trade deficit is good for the economy.

The sheer size and persistence of the U.S. trade deficits and inflows of foreign capital since the 1980s are a legitimate cause for concern. The huge U.S. economy will not be destabilized by an outflow of international capital as easily as, say, the comparatively tiny economies of Thailand and Indonesia were in 1997–1998. Even an economy that is not knocked down, however, can still be shaken. American policymakers should certainly be paying attention to those cases where a pattern of extensive and sustained current account deficits and foreign borrowing has gone badly—if only as a cautionary tale.

**CLEAR IT UP****Are trade surpluses always beneficial? Considering Japan since the 1990s.**

Perhaps no economy around the world is better known for its trade surpluses than Japan. Since 1990, the size of these surpluses has often been near \$100 billion per year. When Japan’s economy was growing vigorously in the 1960s and 1970s, many, especially non-economists, described its large trade surpluses either a cause or a result of its robust economic health. However, from a standpoint of economic growth, Japan’s economy has been teetering in and out of recession since 1990, with real GDP growth averaging only about 1% per year, and an unemployment rate

that has been creeping higher. Clearly, a whopping trade surplus is no guarantee of economic good health.

Instead, Japan's trade surplus reflects that Japan has a very high rate of domestic savings, more than the Japanese economy can invest domestically, and so it invests the extra funds abroad. In Japan's slow economy, consumption of imports is relatively low, and the growth of consumption is relatively slow. Thus, Japan's exports continually exceed its imports, leaving the trade surplus continually high. Recently, Japan's trade surpluses began to deteriorate. In 2013, Japan ran a trade deficit due to the high cost of imported oil. By 2015, Japan again had a surplus and continues to run one today.

23.6 The Difference between Level of Trade and the Trade Balance

LEARNING OBJECTIVES

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

- Identify three factors that influence a country's level of trade
- Differentiate between balance of trade and level of trade

A nation's *level* of trade may at first sound like much the same issue as the *balance* of trade, but these two are actually quite separate. It is perfectly possible for a country to have a very high level of trade—measured by its exports of goods and services as a share of its GDP—while it also has a near-balance between exports and imports. A high level of trade indicates that the nation exports a good portion of its production. It is also possible for a country's trade to be a relatively low share of GDP, relative to global averages, but for the imbalance between its exports and its imports to be quite large. We emphasized this general theme earlier in [Measuring Trade Balances](#), which offered some illustrative figures on trade levels and balances.

A country's level of trade tells how much of its production it exports. We measure this by the percent of exports out of GDP. It indicates the degree of an economy's globalization. Some countries, such as Germany, have a high level of trade—they export almost 50% of their total production. The balance of trade tells us if the country is running a trade surplus or trade deficit. A country can have a low level of trade but a high trade deficit. (For example, the United States only exports around 10% of its GDP, but it has a trade deficit of over \$600 billion.)

Three factors strongly influence a nation's level of trade: the size of its economy, its geographic location, and its history of trade. Large economies like the United States can do much of their trading internally, while small economies like Sweden have less ability to provide what they want internally and tend to have higher ratios of exports and imports to GDP. Nations that are neighbors tend to trade more, since costs of transportation and communication are lower. Moreover, some nations have long and established patterns of international trade, while others do not.

Consequently, a relatively small economy like Sweden, with many nearby trading partners across Europe and a long history of foreign trade, has a high level of trade. Brazil and India, which are fairly large economies that have often sought to inhibit trade in recent decades, have lower levels of trade; whereas, the United States and Japan are extremely large economies that have comparatively few nearby trading partners. Both countries actually have quite low levels of trade by world standards. The ratio of exports to GDP in either the United States or in Japan is about half of the world average.

The balance of trade is a separate issue from the level of trade. The United States has a low level of trade, but had enormous trade deficits for most years from the mid-1980s into the 2000s. Japan has a low level of trade by world standards, but has typically shown large trade surpluses in recent decades. Nations like Germany and the United Kingdom have medium to high levels of trade by world standards, but Germany had a moderate trade surplus in 2020, while the United Kingdom had a moderate trade deficit. Their trade picture was roughly in balance in the late 1990s. Sweden had a high level of trade and a moderate trade surplus in 2020, while Canada had a high level of trade and a moderate trade deficit that same year.

In short, it is quite possible for nations with a relatively low level of trade, expressed as a percentage of GDP, to

have relatively large trade deficits. It is also quite possible for nations with a near balance between exports and imports to worry about the consequences of high levels of trade for the economy. It is not inconsistent to believe that a high level of trade is potentially beneficial to an economy, because of the way it allows nations to play to their comparative advantages, and to also be concerned about any macroeconomic instability caused by a long-term pattern of large trade deficits. The following Clear It Up feature discusses how this sort of dynamic played out in Colonial India.



CLEAR IT UP

Are trade surpluses always beneficial? Considering Colonial India.

India was formally under British rule from 1858 to 1947. During that time, India consistently had trade surpluses with Great Britain. Anyone who believes that trade surpluses are a sign of economic strength and dominance while trade deficits are a sign of economic weakness must find this pattern odd, since it would mean that colonial India was successfully dominating and exploiting Great Britain for almost a century—which was not true.

Instead, India's trade surpluses with Great Britain meant that each year there was an overall flow of financial capital from India to Great Britain. In India, many heavily criticized this financial capital flow as the “drain,” and they viewed eliminating the financial capital drain as one of the many reasons why India would benefit from achieving independence.

Final Thoughts about Trade Balances

Trade deficits can be a good or a bad sign for an economy, and trade surpluses can be a good or a bad sign. Even a trade balance of zero—which just means that a nation is neither a net borrower nor lender in the international economy—can be either a good or bad sign. The fundamental economic question is not whether a nation's economy is borrowing or lending at all, but whether the particular borrowing or lending in the particular economic conditions of that country makes sense.

It is interesting to reflect on how public attitudes toward trade deficits and surpluses might change if we could somehow change the labels that people and the news media affix to them. If we called a trade deficit “attracting foreign financial capital”—which accurately describes what a trade deficit means—then trade deficits might look more attractive. Conversely, if we called a trade surplus “shipping financial capital abroad”—which accurately captures what a trade surplus does—then trade surpluses might look less attractive. Either way, the key to understanding trade balances is to understand the relationships between flows of trade and flows of international payments, and what these relationships imply about the causes, benefits, and risks of different kinds of trade balances. The first step along this journey of understanding is to move beyond knee-jerk reactions to terms like “trade surplus,” “trade balance,” and “trade deficit.”



BRING IT HOME

More than Meets the Eye in the Congo

Now that you see the big picture, you undoubtedly realize that all of the economic choices you make, such as depositing savings or investing in an international mutual fund, do influence the flow of goods and services as well as the flows of money around the world.

You now know that a trade surplus does not necessarily tell us whether an economy is performing well or not. The Democratic Republic of the Congo ran a trade surplus in 2013, as we learned in the beginning of the chapter. Yet its current account balance was $-\$2.8$ billion. However, the return of political stability and the rebuilding in the aftermath of the civil war there has meant a flow of investment and financial capital into the country. In this case, a negative current account balance means the country is being rebuilt—and that is a good thing.

Key Terms

balance of trade (trade balance) the gap, if any, between a nation's exports and imports

current account balance a broad measure of the balance of trade that includes trade in goods and services, as well as international flows of income and foreign aid

exports of goods and services as a percentage of GDP the dollar value of exports divided by the dollar value of a country's GDP

financial capital the international flows of money that facilitates trade and investment

merchandise trade balance the balance of trade looking only at goods

national savings and investment identity the total of private savings and public savings (a government budget surplus)

unilateral transfers "one-way payments" that governments, private entities, or individuals make that they sent abroad with nothing received in return

Key Concepts and Summary

23.1 Measuring Trade Balances

The trade balance measures the gap between a country's exports and its imports. In most high-income economies, goods comprise less than half of a country's total production, while services comprise more than half. The last two decades have seen a surge in international trade in services; however, most global trade still takes the form of goods rather than services. The current account balance includes the trade in goods, services, and money flowing into and out of a country from investments and unilateral transfers.

23.2 Trade Balances in Historical and International Context

The United States developed large trade surpluses in the early 1980s, swung back to a tiny trade surplus in 1991, and then had even larger trade deficits in the late 1990s and early 2000s. As we will see below, a trade deficit necessarily means a net inflow of financial capital from abroad, while a trade surplus necessarily means a net outflow of financial capital from an economy to other countries.

23.3 Trade Balances and Flows of Financial Capital

International flows of goods and services are closely connected to the international flows of financial capital. A current account deficit means that, after taking all the flows of payments from goods, services, and income together, the country is a net borrower from the rest of the world. A current account surplus is the opposite and means the country is a net lender to the rest of the world.

23.4 The National Saving and Investment Identity

The national saving and investment identity is based on the relationship that the total quantity of financial capital supplied from all sources must equal the total quantity of financial capital demanded from all sources. If S is private saving, T is taxes, G is government spending, M is imports, X is exports, and I is investment, then for an economy with a current account deficit and a budget deficit:

$$\text{Supply of financial capital} = \text{Demand for financial capital}$$

$$S + (M - X) = I + (G - T)$$

A recession tends to increase the trade balance (meaning a higher trade surplus or lower trade deficit), while economic boom will tend to decrease the trade balance (meaning a lower trade surplus or a larger trade deficit).

23.5 The Pros and Cons of Trade Deficits and Surpluses

Trade surpluses are no guarantee of economic health, and trade deficits are no guarantee of economic weakness. Either trade deficits or trade surpluses can work out well or poorly, depending on whether a government wisely invests the corresponding flows of financial capital.

23.6 The Difference between Level of Trade and the Trade Balance

There is a difference between the level of a country's trade and the balance of trade. The government measures its level of trade by the percentage of exports out of GDP, or the size of the economy. Small economies that have nearby trading partners and a history of international trade will tend to have higher levels of trade. Larger economies with few nearby trading partners and a limited history of international trade will tend to have lower levels of trade. The level of trade is different from the trade balance. The level of trade depends on a country's history of trade, its geography, and the size of its economy. A country's balance of trade is the dollar difference between its exports and imports.

Trade deficits and trade surpluses are not necessarily good or bad—it depends on the circumstances. Even if a country is borrowing, if it invests that money in productivity-boosting investments it can lead to an improvement in long-term economic growth.

Self-Check Questions

1. If foreign investors buy more U.S. stocks and bonds, how would that show up in the current account balance?
2. If the trade deficit of the United States increases, how is the current account balance affected?
3. State whether each of the following events involves a financial flow to the Mexican economy or a financial flow out of the Mexican economy:
 - a. Mexico imports services from Japan
 - b. Mexico exports goods to Canada
 - c. U.S. investors receive a return from past financial investments in Mexico
4. In what way does comparing a country's exports to GDP reflect its degree of globalization?
5. At one point Canada's GDP was \$1,800 billion and its exports were \$542 billion. What was Canada's export ratio at this time?
6. The GDP for the United States is \$18,036 billion and its current account balance is −\$484 billion. What percent of GDP is the current account balance?
7. Why does the trade balance and the current account balance track so closely together over time?
8. State whether each of the following events involves a financial flow to the U.S. economy or away from the U.S. economy:
 - a. Export sales to Germany
 - b. Returns paid on past U.S. financial investments in Brazil
 - c. Foreign aid from the U.S. government to Egypt
 - d. Imported oil from the Russian Federation
 - e. Japanese investors buying U.S. real estate
9. How does the bottom portion of [Figure 23.3](#), showing the international flow of investments and capital, differ from the upper portion?
10. Explain the relationship between a current account deficit or surplus and the flow of funds.
11. Using the national savings and investment identity, explain how each of the following changes (*ceteris paribus*) will increase or decrease the trade balance:
 - a. A lower domestic savings rate
 - b. The government changes from running a budget surplus to running a budget deficit
 - c. The rate of domestic investment surges

12. If a country is running a government budget surplus, why is $(T - G)$ on the left side of the saving-investment identity?
13. What determines the size of a country's trade deficit?
14. If domestic investment increases, and there is no change in the amount of private and public saving, what must happen to the size of the trade deficit?
15. Why does a recession cause a trade deficit to increase?
16. Both the United States and global economies are booming. Will U.S. imports and/or exports increase?
17. For each of the following, indicate which type of government spending would justify a budget deficit and which would not.
 - a. Increased federal spending on Medicare
 - b. Increased spending on education
 - c. Increased spending on the space program
 - d. Increased spending on airports and air traffic control
18. How did large trade deficits hurt the East Asian countries in the mid 1980s? (Recall that trade deficits are equivalent to inflows of financial capital from abroad.)
19. Describe a scenario in which a trade surplus benefits an economy and one in which a trade surplus is occurring in an economy that performs poorly. What key factor or factors are making the difference in the outcome that results from a trade surplus?
20. The United States exports 14% of GDP while Germany exports about 50% of its GDP. Explain what that means.
21. Explain briefly whether each of the following would be more likely to lead to a higher level of trade for an economy, or a greater imbalance of trade for an economy.
 - a. Living in an especially large country
 - b. Having a domestic investment rate much higher than the domestic savings rate
 - c. Having many other large economies geographically nearby
 - d. Having an especially large budget deficit
 - e. Having countries with a tradition of strong protectionist legislation shutting out imports

Review Questions

22. If imports exceed exports, is it a trade deficit or a trade surplus? What about if exports exceed imports?
23. What is included in the current account balance?
24. In recent decades, has the U.S. trade balance usually been in deficit, surplus, or balanced?
25. Does a trade surplus mean an overall inflow of financial capital to an economy, or an overall outflow of financial capital? What about a trade deficit?
26. What are the two main sides of the national savings and investment identity?
27. What are the main components of the national savings and investment identity?
28. When is a trade deficit likely to work out well for an economy? When is it likely to work out poorly?
29. Does a trade surplus help to guarantee strong economic growth?
30. What three factors will determine whether a nation has a higher or lower share of trade relative to its GDP?
31. What is the difference between trade deficits and balance of trade?

Critical Thinking Questions

32. Occasionally, a government official will argue that a country should strive for both a trade surplus and a healthy inflow of capital from abroad. Explain why such a statement is economically impossible.
33. A government official announces a new policy. The country wishes to eliminate its trade deficit, but will strongly encourage financial investment from foreign firms. Explain why such a statement is contradictory.
34. If a country is a big exporter, is it more exposed to global financial crises?
35. If countries reduced trade barriers, would the international flows of money increase?
36. Is it better for your country to be an international lender or borrower?
37. Many think that the size of a trade deficit is due to a lack of competitiveness of domestic sectors, such as autos. Explain why this is not true.
38. If you observed a country with a rapidly growing trade surplus over a period of a year or so, would you be more likely to believe that the country's economy was in a period of recession or of rapid growth? Explain.
39. Occasionally, a government official will argue that a country should strive for both a trade surplus and a healthy inflow of capital from abroad. Is this possible?
40. What is more important, a country's current account balance or GDP growth? Why?
41. Will nations that are more involved in foreign trade tend to have higher trade imbalances, lower trade imbalances, or is the pattern unpredictable?
42. Some economists warn that the persistent trade deficits and a negative current account balance that the United States has run will be a problem in the long run. Do you agree or not? Explain your answer.

Problems

43. In 2001, the United Kingdom's economy exported goods worth £192 billion and services worth another £77 billion. It imported goods worth £225 billion and services worth £66 billion. Receipts of income from abroad were £140 billion while income payments going abroad were £131 billion. Government transfers from the United Kingdom to the rest of the world were £23 billion, while various U.K. government agencies received payments of £16 billion from the rest of the world.
 - a. Calculate the U.K. merchandise trade deficit for 2001.
 - b. Calculate the current account balance for 2001.
 - c. Explain how you decided whether payments on foreign investment and government transfers counted on the positive or the negative side of the current account balance for the United Kingdom in 2001.
44. Imagine that the U.S. economy finds itself in the following situation: a government budget deficit of \$100 billion, total domestic savings of \$1,500 billion, and total domestic physical capital investment of \$1,600 billion. According to the national saving and investment identity, what will be the current account balance? What will be the current account balance if investment rises by \$50 billion, while the budget deficit and national savings remain the same?

45. [Table 23.7](#) provides some hypothetical data on macroeconomic accounts for three countries represented by A, B, and C and measured in billions of currency units. In [Table 23.7](#), private household saving is SH, tax revenue is T, government spending is G, and investment spending is I.

	A	B	C
SH	700	500	600
T	00	500	500
G	600	350	650
I	800	400	450

TABLE 23.7
Macroeconomic
Accounts

- Calculate the trade balance and the net inflow of foreign saving for each country.
 - State whether each one has a trade surplus or deficit (or balanced trade).
 - State whether each is a net lender or borrower internationally and explain.
46. Imagine that the economy of Germany finds itself in the following situation: the government budget has a surplus of 1% of Germany's GDP; private savings is 20% of GDP; and physical investment is 18% of GDP.
- Based on the national saving and investment identity, what is the current account balance?
 - If the government budget surplus falls to zero, how will this affect the current account balance?

The Aggregate Demand/Aggregate Supply Model

24



FIGURE 24.1 New Home Construction At the peak of the housing bubble, many people across the country were able to secure the loans necessary to build new houses. (Credit: modification of "our house! Again!" by Tim Pierce/ Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Macroeconomic Perspectives on Demand and Supply
- Building a Model of Aggregate Supply and Aggregate Demand
- Shifts in Aggregate Supply
- Shifts in Aggregate Demand
- How the AS–AD Model Incorporates Growth, Unemployment, and Inflation
- Keynes' Law and Say's Law in the AS–AD Model

Introduction to the Aggregate Supply–Aggregate Demand Model



BRING IT HOME

From Housing Bubble to Housing Bust

The United States experienced rising home ownership rates for most of the last two decades. Between 1990 and 2006, the U.S. housing market grew. Homeownership rates grew from 64% to a high of over 69% between 2004 and 2005. For many people, this was a period in which they could either buy first homes or buy a larger and more expensive home. During this time mortgage values tripled. Housing became more accessible to Americans and was considered to be a safe financial investment. [Figure 24.2](#) shows how new single family home sales peaked in 2005 at 1,279,000 units.

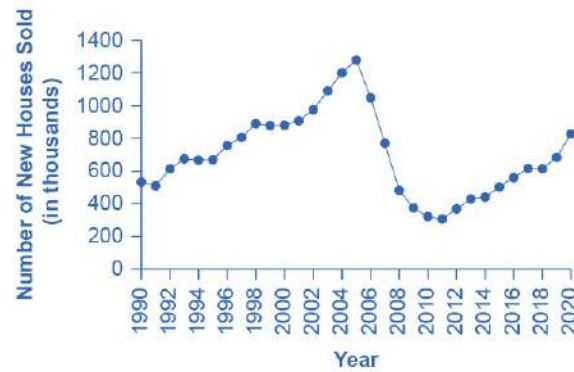


FIGURE 24.2 New Single Family Houses Sold From the early 1990s up through 2005, the number of new single family houses sold rose steadily. In 2006, the number dropped dramatically and this dramatic decline continued through 2011. Beginning in 2012, the number of new houses sold began to climb back up. (Source: U.S. Census Bureau)

The housing bubble began to show signs of bursting in 2005, as delinquency and late payments began to grow and an oversupply of new homes on the market became apparent. Dropping home values contributed to a decrease in the overall wealth of the household sector and caused homeowners to pull back on spending. Several mortgage lenders were forced to file for bankruptcy because homeowners were not making their payments, and by 2008 the problem had spread throughout the financial markets. Lenders clamped down on credit and the housing bubble burst. Financial markets were now in crisis and unable or unwilling to even extend credit to credit-worthy customers.

The housing bubble and the crisis in the financial markets were major contributors to the Great Recession that led to unemployment rates over 10% and falling GDP. While the United States is still recovering from the impact of the Great Recession, it has made substantial progress in restoring financial market stability through implementing aggressive fiscal and monetary policy.

The economic history of the United States is cyclical in nature with recessions and expansions. Some of these fluctuations are severe, such as the economic downturn that occurred during the Great Depression in the 1930s which lasted several years. Why does the economy grow at different rates in different years? What are the causes of the cyclical behavior of the economy? This chapter will introduce an important model, the aggregate demand–aggregate supply model, to begin our understanding of why economies expand and contract over time.

[Click to view content \(https://openstax.org/books/principles-economics-3e/pages/24-introduction-to-the-aggregate-supplyaggregate-demand-model\)](https://openstax.org/books/principles-economics-3e/pages/24-introduction-to-the-aggregate-supplyaggregate-demand-model)

New One-Family Houses Sold in the United States.

A key part of macroeconomics is the use of models to analyze macro issues and problems. How is the rate of economic growth connected to changes in the unemployment rate? Is there a reason why unemployment and inflation seem to move in opposite directions: lower unemployment and higher inflation from 1997 to 2000, higher unemployment and lower inflation in the early 2000s, lower unemployment and higher inflation in the mid-2000s, and then higher unemployment and lower inflation in 2009? Why did the current account deficit rise so high, but then decline in 2009?

To analyze questions like these, we must move beyond discussing macroeconomic issues one at a time, and begin building economic models that will capture the relationships and interconnections between them. The next three chapters take up this task. This chapter introduces the macroeconomic model of aggregate supply and aggregate demand, how the two interact to reach a macroeconomic equilibrium, and how shifts in aggregate demand or aggregate supply will affect that equilibrium. This chapter also relates the model of aggregate supply and aggregate demand to the three goals of economic policy (growth, unemployment, and

inflation), and provides a framework for thinking about many of the connections and tradeoffs between these goals. The chapter on [The Keynesian Perspective](#) focuses on the macroeconomy in the short run, where aggregate demand plays a crucial role. The chapter on [The Neoclassical Perspective](#) explores the macroeconomy in the long run, where aggregate supply plays a crucial role.

24.1 Macroeconomic Perspectives on Demand and Supply

LEARNING OBJECTIVES

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

- Explain Say's Law and understand why it primarily applies in the long run
- Explain Keynes' Law and understand why it primarily applies in the short run

Macroeconomists over the last two centuries have often divided into two groups: those who argue that supply is the most important determinant of the size of the macroeconomy while demand just tags along, and those who argue that demand is the most important factor in the size of the macroeconomy while supply just tags along.

Say's Law and the Macroeconomics of Supply

Those economists who emphasize the role of supply in the macroeconomy often refer to the work of a famous early nineteenth century French economist named Jean-Baptiste Say (1767–1832). **Say's law** is: "Supply creates its own demand." As a matter of historical accuracy, it seems clear that Say never actually wrote down this law and that it oversimplifies his beliefs, but the law lives on as useful shorthand for summarizing a point of view.

The intuition behind Say's law is that each time a good or service is produced and sold, it generates income that is earned for someone: a worker, a manager, an owner, or those who are workers, managers, and owners at firms that supply inputs along the chain of production. We alluded to this earlier in our discussion of the National Income approach to measuring GDP. The forces of supply and demand in individual markets will cause prices to rise and fall. The bottom line remains, however, that every sale represents income to someone, and so, Say's law argues, a given value of supply must create an equivalent value of demand somewhere else in the economy. Because Jean-Baptiste Say, Adam Smith, and other economists writing around the turn of the nineteenth century who discussed this view were known as "classical" economists, modern economists who generally subscribe to the Say's law view on the importance of supply for determining the size of the macroeconomy are called **neoclassical economists**.

If supply always creates exactly enough demand at the macroeconomic level, then (as Say himself recognized) it is hard to understand why periods of recession and high unemployment should ever occur. To be sure, even if total supply always creates an equal amount of total demand, the economy could still experience a situation of some firms earning profits while other firms suffer losses. Nevertheless, a recession is not a situation where all business failures are exactly counterbalanced by an offsetting number of successes. A recession is a situation in which the economy as a whole is shrinking in size, business failures outnumber the remaining success stories, and many firms end up suffering losses and laying off workers.

Say's law that supply creates its own demand does seem a good approximation for the long run. Over periods of some years or decades, as the productive power of an economy to supply goods and services increases, total demand in the economy grows at roughly the same pace. However, over shorter time horizons of a few months or even years, recessions or even depressions occur in which firms, as a group, seem to face a lack of demand for their products.

Keynes' Law and the Macroeconomics of Demand

The alternative to Say's law, with its emphasis on supply, is **Keynes' law**: "Demand creates its own supply." As a matter of historical accuracy, just as Jean-Baptiste Say never wrote down anything as simpleminded as Say's law, John Maynard Keynes never wrote down Keynes' law, but the law is a useful simplification that conveys a

certain point of view.

When Keynes wrote his influential work *The General Theory of Employment, Interest, and Money* during the 1930s Great Depression, he pointed out that during the Depression, the economy's capacity to supply goods and services had not changed much. U.S. unemployment rates soared higher than 20% from 1933 to 1935, but the number of possible workers had not increased or decreased much. Factories closed, but machinery and equipment had not disappeared. Technologies that had been invented in the 1920s were not un-invented and forgotten in the 1930s. Thus, Keynes argued that the Great Depression—and many ordinary recessions as well—were not caused by a drop in the ability of the economy to supply goods as measured by labor, physical capital, or technology. He argued the economy often produced less than its full potential, not because it was technically impossible to produce more with the existing workers and machines, but because a lack of demand in the economy as a whole led to inadequate incentives for firms to produce. In such cases, he argued, the level of GDP in the economy was not primarily determined by the potential of what the economy could supply, but rather by the amount of total demand.

Keynes' law seems to apply fairly well in the short run of a few months to a few years, when many firms experience either a drop in demand for their output during a recession or so much demand that they have trouble producing enough during an economic boom. However, demand cannot tell the whole macroeconomic story, either. After all, if demand was all that mattered at the macroeconomic level, then the government could make the economy as large as it wanted just by pumping up total demand through a large increase in the government spending component or by legislating large tax cuts to push up the consumption component. Economies do, however, face genuine limits to how much they can produce, limits determined by the quantity of labor, physical capital, technology, and the institutional and market structures that bring these factors of production together. These constraints on what an economy can supply at the macroeconomic level do not disappear just because of an increase in demand.

Combining Supply and Demand in Macroeconomics

Two insights emerge from this overview of Say's law with its emphasis on macroeconomic supply and Keynes' law with its emphasis on macroeconomic demand. The first conclusion, which is not exactly a hot news flash, is that an economic approach focused only on the supply side or only on the demand side can be only a partial success. We need to take into account both supply and demand. The second conclusion is that since Keynes' law applies more accurately in the short run and Say's law applies more accurately in the long run, the tradeoffs and connections between the three goals of macroeconomics may be different in the short run and the long run.

24.2 Building a Model of Aggregate Demand and Aggregate Supply

LEARNING OBJECTIVES

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

- Explain the aggregate supply curve and how it relates to real GDP and potential GDP
- Explain the aggregate demand curve and how it is influenced by price levels
- Interpret the aggregate demand/aggregate supply model
- Identify the point of equilibrium in the aggregate demand/aggregate supply model
- Define short run aggregate supply and long run aggregate supply

To build a useful macroeconomic model, we need a model that shows what determines total supply or total demand for the economy, and how total demand and total supply interact at the macroeconomic level. We call this the **aggregate demand/aggregate supply model**. This module will explain aggregate supply, aggregate demand, and the equilibrium between them. The following modules will discuss the causes of shifts in aggregate supply and aggregate demand.

The Aggregate Supply Curve and Potential GDP

Firms make decisions about what quantity to supply based on the profits they expect to earn. They determine profits, in turn, by the price of the outputs they sell and by the prices of the inputs, like labor or raw materials, that they need to buy. **Aggregate supply (AS)** refers to the total quantity of output (i.e. real GDP) firms will produce and sell. The **aggregate supply (AS) curve** shows the total quantity of output (i.e. real GDP) that firms will produce and sell at each price level.

Figure 24.3 shows an aggregate supply curve. In the following paragraphs, we will walk through the elements of the diagram one at a time: the horizontal and vertical axes, the aggregate supply curve itself, and the meaning of the potential GDP vertical line.

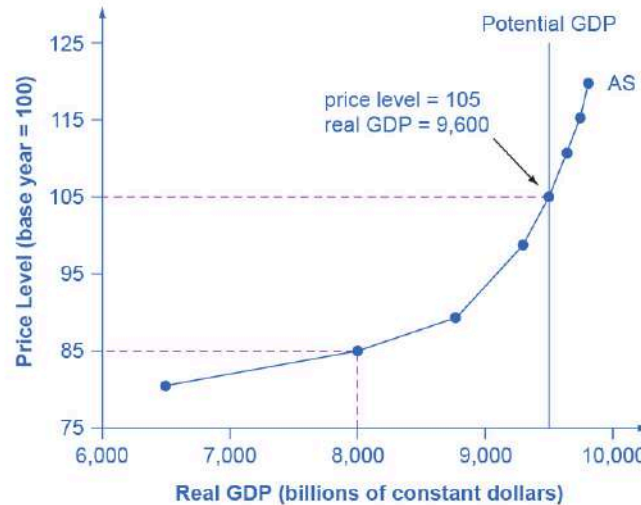


FIGURE 24.3 The Aggregate Supply Curve Aggregate supply (AS) slopes up, because as the price level for outputs rises, with the price of inputs remaining fixed, firms have an incentive to produce more to earn higher profits. The potential GDP line shows the maximum that the economy can produce with full employment of workers and physical capital.

The diagram's horizontal axis shows real GDP—that is, the level of GDP adjusted for inflation. The vertical axis shows the price level, which measures the average price of all goods and services produced in the economy. In other words, the price level in the AD-AS model is what we called the GDP Deflator in [The Macroeconomic Perspective](#). Remember that the price level is different from the inflation rate. Visualize the price level as an index number, like the Consumer Price Index, while the inflation rate is the percentage change in the price level over time.

As the price level rises, real GDP rises as well. Why? The price level on the vertical axis represents prices for final goods or outputs bought in the economy—i.e. the GDP deflator—not the price level for intermediate goods and services that are inputs to production. Thus, the AS curve describes how suppliers will react to a higher price level for final outputs of goods and services, while holding the prices of inputs like labor and energy constant. If firms across the economy face a situation where the price level of what they produce and sell is rising, but their costs of production are not rising, then the lure of higher profits will induce them to expand production. In other words, an aggregate supply curve shows how producers as a group will respond to an increase in aggregate demand.

An AS curve's slope changes from nearly flat at its far left to nearly vertical at its far right. At the far left of the aggregate supply curve, the level of output in the economy is far below **potential GDP**, which we define as the amount of real GDP an economy can produce by fully employing its existing levels of labor, physical capital, and technology, in the context of its existing market and legal institutions. At these relatively low levels of output, levels of unemployment are high, and many factories are running only part-time, or have closed their

doors. In this situation, a relatively small increase in the prices of the outputs that businesses sell—while assuming no rise in input prices—can encourage a considerable surge in the quantity of aggregate supply because so many workers and factories are ready to swing into production.

As the GDP increases, however, some firms and industries will start running into limits: perhaps nearly all of the expert workers in a certain industry will have jobs or factories in certain geographic areas or industries will be running at full speed. In the AS curve's intermediate area, a higher price level for outputs continues to encourage a greater quantity of output—but as the increasingly steep upward slope of the aggregate supply curve shows, the increase in real GDP in response to a given rise in the price level will not be as large. (Read the following Clear It Up feature to learn why the AS curve crosses potential GDP.)



CLEAR IT UP

Why does AS cross potential GDP?

Economists typically draw the aggregate supply curve to cross the potential GDP line. This shape may seem puzzling: How can an economy produce at an output level which is higher than its “potential” or “full employment” GDP? The economic intuition here is that if prices for outputs were high enough, producers would make fanatical efforts to produce: all workers would be on double-overtime, all machines would run 24 hours a day, seven days a week. Such hyper-intense production would go beyond using potential labor and physical capital resources fully, to using them in a way that is not sustainable in the long term. Thus, it is possible for production to sprint above potential GDP, but only in the short run.

At the far right, the aggregate supply curve becomes nearly vertical. At this quantity, higher prices for outputs cannot encourage additional output, because even if firms want to expand output, the inputs of labor and machinery in the economy are fully employed. In this example, the vertical line in the exhibit shows that potential GDP occurs at a total output of 9,500. When an economy is operating at its potential GDP, machines and factories are running at capacity, and the unemployment rate is relatively low—at the natural rate of unemployment. For this reason, potential GDP is sometimes also called **full-employment GDP**.

The Aggregate Demand Curve

Aggregate demand (AD) refers to the amount of total spending on domestic goods and services in an economy. (Strictly speaking, AD is what economists call total planned expenditure. We will further explain this distinction in the appendix [The Expenditure-Output Model](#). For now, just think of aggregate demand as total spending.) It includes all four components of demand: consumption, investment, government spending, and net exports (exports minus imports). This demand is determined by a number of factors, but one of them is the price level—recall though, that the price level is an index number such as the GDP deflator that measures the average price of the things we buy. The **aggregate demand (AD) curve** shows the total spending on domestic goods and services at each price level.

[Figure 24.4](#) presents an aggregate demand (AD) curve. Just like the aggregate supply curve, the horizontal axis shows real GDP and the vertical axis shows the price level. The AD curve slopes down, which means that increases in the price level of outputs lead to a lower quantity of total spending. The reasons behind this shape are related to how changes in the price level affect the different components of aggregate demand. The following components comprise aggregate demand: consumption spending (C), investment spending (I), government spending (G), and spending on exports (X) minus imports (M): $C + I + G + X - M$.

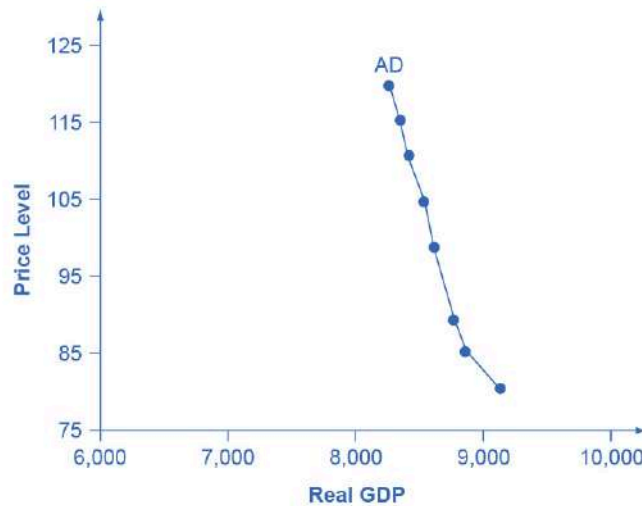


FIGURE 24.4 The Aggregate Demand Curve Aggregate demand (AD) slopes down, showing that, as the price level rises, the amount of total spending on domestic goods and services declines.

The wealth effect holds that as the price level increases, the buying power of savings that people have stored up in bank accounts and other assets will diminish, eaten away to some extent by inflation. Because a rise in the price level reduces people's wealth, consumption spending will fall as the price level rises.

The interest rate effect is that as prices for outputs rise, the same purchases will take more money or credit to accomplish. This additional demand for money and credit will push interest rates higher. In turn, higher interest rates will reduce borrowing by businesses for investment purposes and reduce borrowing by households for homes and cars—thus reducing consumption and investment spending.

The foreign price effect points out that if prices rise in the United States while remaining fixed in other countries, then goods in the United States will be relatively more expensive compared to goods in the rest of the world. U.S. exports will be relatively more expensive, and the quantity of exports sold will fall. U.S. imports from abroad will be relatively cheaper, so the quantity of imports will rise. Thus, a higher domestic price level, relative to price levels in other countries, will reduce net export expenditures.

Among economists all three of these effects are controversial, in part because they do not seem to be very large. For this reason, the aggregate demand curve in [Figure 24.4](#) slopes downward fairly steeply. The steep slope indicates that a higher price level for final outputs reduces aggregate demand for all three of these reasons, but that the change in the quantity of aggregate demand as a result of changes in price level is not very large.

Read the following Work It Out feature to learn how to interpret the AD/AS model. In this example, aggregate supply, aggregate demand, and the price level are given for the imaginary country of Xurbia.

WORK IT OUT

Interpreting the AD/AS Model

[Table 24.1](#) shows information on aggregate supply, aggregate demand, and the price level for the imaginary country of Xurbia. What information does [Table 24.1](#) tell you about the state of the Xurbia's economy? Where is the equilibrium price level and output level (this is the SR macroequilibrium)? Is Xurbia risking inflationary pressures or facing high unemployment? How can you tell?

Price Level	Aggregate Demand	Aggregate Supply
110	\$700	\$600
120	\$690	\$640
130	\$680	\$680
140	\$670	\$720
150	\$660	\$740
160	\$650	\$760
170	\$640	\$770

TABLE 24.1 Price Level: Aggregate Demand/Aggregate Supply

To begin to use the AD/AS model, it is important to plot the AS and AD curves from the data provided. What is the equilibrium?

Step 1. Draw your x- and y-axis. Label the x-axis Real GDP and the y-axis Price Level.

Step 2. Plot AD on your graph.

Step 3. Plot AS on your graph.

Step 4. Look at [Figure 24.5](#) which provides a visual to aid in your analysis.

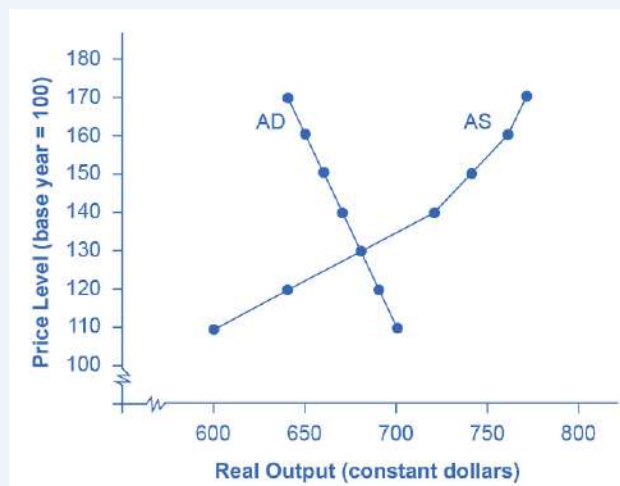


FIGURE 24.5 The AD/AS Curves AD and AS curves created from the data in [Table 24.1](#).

Step 5. Determine where AD and AS intersect. This is the equilibrium with price level at 130 and real GDP at \$680.

Step 6. Look at the graph to determine where equilibrium is located. We can see that this equilibrium is fairly far from where the AS curve becomes near-vertical (or at least quite steep) which seems to start at about \$750 of real output. This implies that the economy is not close to potential GDP. Thus, unemployment will be high. In the relatively flat part of the AS curve, where the equilibrium occurs, changes in the price level will not be a major concern, since such changes are likely to be small.

Step 7. Determine what the steep portion of the AS curve indicates. Where the AS curve is steep, the economy is at or close to potential GDP.

Step 8. Draw conclusions from the given information:

- If equilibrium occurs in the flat range of AS, then economy is not close to potential GDP and will be experiencing unemployment, but stable price level.
- If equilibrium occurs in the steep range of AS, then the economy is close or at potential GDP and will be experiencing rising price levels or inflationary pressures, but will have a low unemployment rate.

Equilibrium in the Aggregate Demand/Aggregate Supply Model

The intersection of the aggregate supply and aggregate demand curves shows the equilibrium level of real GDP and the equilibrium price level in the economy. At a relatively low price level for output, firms have little incentive to produce, although consumers would be willing to purchase a large quantity of output. As the price level rises, aggregate supply rises and aggregate demand falls until the equilibrium point is reached.

Figure 24.6 combines the AS curve from Figure 24.3 and the AD curve from Figure 24.4 and places them both on a single diagram. In this example, the equilibrium point occurs at point E, at a price level of 90 and an output level of 8,800.

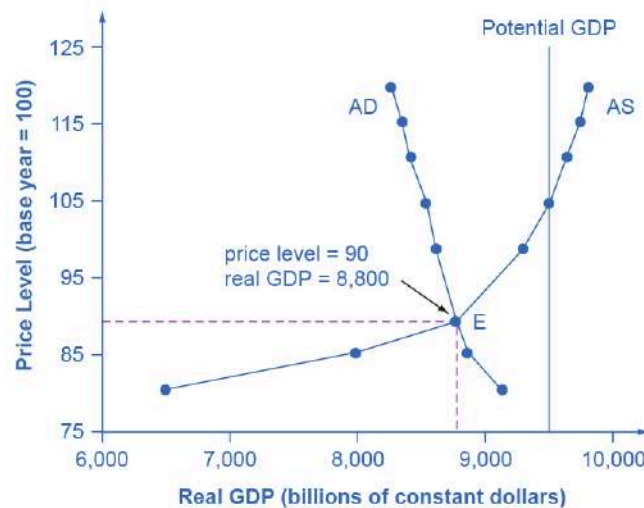


FIGURE 24.6 Aggregate Supply and Aggregate Demand The equilibrium, where aggregate supply (AS) equals aggregate demand (AD), occurs at a price level of 90 and an output level of 8,800.

Confusion sometimes arises between the aggregate supply and aggregate demand model and the microeconomic analysis of demand and supply in particular markets for goods, services, labor, and capital. Read the following Clear It Up feature to gain an understanding of whether AS and AD are macro or micro.



CLEAR IT UP

Are AS and AD macro or micro?

These aggregate supply and demand models and the microeconomic analysis of demand and supply in particular markets for goods, services, labor, and capital have a superficial resemblance, but they also have many underlying differences.

For example, the vertical and horizontal axes have distinctly different meanings in macroeconomic and

microeconomic diagrams. The vertical axis of a microeconomic demand and supply diagram expresses a price (or wage or rate of return) for an individual good or service. This price is implicitly relative: it is intended to be compared with the prices of other products (for example, the price of pizza relative to the price of fried chicken). In contrast, the vertical axis of an aggregate supply and aggregate demand diagram expresses the level of a price index like the Consumer Price Index or the GDP deflator—combining a wide array of prices from across the economy. The price level is absolute: it is not intended to be compared to any other prices since it is essentially the average price of all products in an economy. The horizontal axis of a microeconomic supply and demand curve measures the quantity of a particular good or service. In contrast, the horizontal axis of the aggregate demand and aggregate supply diagram measures GDP, which is the sum of all the final goods and services produced in the economy, not the quantity in a specific market.

In addition, the economic reasons for the shapes of the curves in the macroeconomic model are different from the reasons behind the shapes of the curves in microeconomic models. Demand curves for individual goods or services slope down primarily because of the existence of substitute goods, not the wealth effects, interest rate, and foreign price effects associated with aggregate demand curves. The slopes of individual supply and demand curves can have a variety of different slopes, depending on the extent to which quantity demanded and quantity supplied react to price in that specific market, but the slopes of the AS and AD curves are much the same in every diagram (although as we shall see in later chapters, short-run and long-run perspectives will emphasize different parts of the AS curve).

In short, just because the AD/AS diagram has two lines that cross, do not assume that it is the same as every other diagram where two lines cross. The intuitions and meanings of the macro and micro diagrams are only distant cousins from different branches of the economics family tree.

Defining SRAS and LRAS

In the Clear It Up feature titled “Why does AS cross potential GDP?” we differentiated between short run changes in aggregate supply which the AS curve shows and long run changes in aggregate supply which the vertical line at potential GDP defines. In the short run, if demand is too low (or too high), it is possible for producers to supply less GDP (or more GDP) than potential. In the long run, however, producers are limited to producing at potential GDP. For this reason, we may also refer to what we have been calling the AS curve as the **short run aggregate supply (SRAS) curve**. We may also refer to the vertical line at potential GDP as the **long run aggregate supply (LRAS) curve**.

24.3 Shifts in Aggregate Supply

LEARNING OBJECTIVES

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

- Explain how productivity growth changes the aggregate supply curve
- Explain how changes in input prices change the aggregate supply curve

The original equilibrium in the AD/AS diagram will shift to a new equilibrium if the AS or AD curve shifts. When the aggregate supply curve shifts to the right, then at every price level, producers supply a greater quantity of real GDP. When the AS curve shifts to the left, then at every price level, producers supply a lower quantity of real GDP. This module discusses two of the most important factors that can lead to shifts in the AS curve: productivity growth and changes in input prices.

How Productivity Growth Shifts the AS Curve

In the long run, the most important factor shifting the AS curve is productivity growth. Productivity means how much output can be produced with a given quantity of labor. One measure of this is output per worker or GDP per capita. Over time, productivity grows so that the same quantity of labor can produce more output. Historically, the real growth in GDP per capita in an advanced economy like the United States has averaged

about 2% to 3% per year, but productivity growth has been faster during certain extended periods like the 1960s and the late 1990s through the early 2000s, or slower during periods like the 1970s. A higher level of productivity shifts the AS curve to the right, because with improved productivity, firms can produce a greater quantity of output at every price level. [Figure 24.7](#) (a) shows an outward shift in productivity over two time periods. The AS curve shifts out from $SRAS_0$ to $SRAS_1$ to $SRAS_2$, and the equilibrium shifts from E_0 to E_1 to E_2 . Note that with increased productivity, workers can produce more GDP. Thus, full employment corresponds to a higher level of potential GDP, which we show as a rightward shift in LRAS from $LRAS_0$ to $LRAS_1$ to $LRAS_2$.

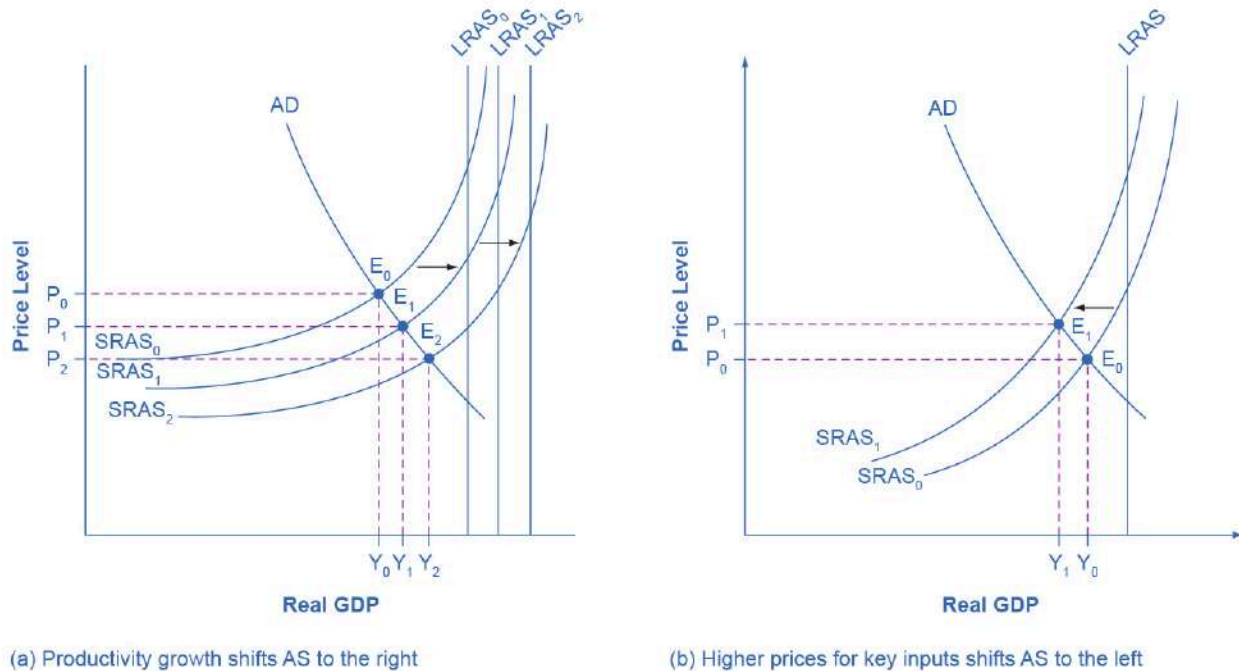


FIGURE 24.7 Shifts in Aggregate Supply (a) The rise in productivity causes the SRAS curve to shift to the right. The original equilibrium E_0 is at the intersection of AD and $SRAS_0$. When SRAS shifts right, then the new equilibrium E_1 is at the intersection of AD and $SRAS_1$, and then yet another equilibrium, E_2 , is at the intersection of AD and $SRAS_2$. Shifts in SRAS to the right, lead to a greater level of output and to downward pressure on the price level. (b) A higher price for inputs means that at any given price level for outputs, a lower real GDP will be produced so aggregate supply will shift to the left from $SRAS_0$ to $SRAS_1$. The new equilibrium, E_1 , has a reduced quantity of output and a higher price level than the original equilibrium (E_0).

A shift in the SRAS curve to the right will result in a greater real GDP and downward pressure on the price level, if aggregate demand remains unchanged. However, if this shift in SRAS results from gains in productivity growth, which we typically measure in terms of a few percentage points per year, the effect will be relatively small over a few months or even a couple of years. Recall how in [Choice in a World of Scarcity](#), we said that a nation's production possibilities frontier is fixed in the short run, but shifts out in the long run? This is the same phenomenon using a different model.

How Changes in Input Prices Shift the AS Curve

Higher prices for inputs that are widely used across the entire economy can have a macroeconomic impact on aggregate supply. Examples of such widely used inputs include labor and energy products. Increases in the price of such inputs will cause the SRAS curve to shift to the left, which means that at each given price level for outputs, a higher price for inputs will discourage production because it will reduce the possibilities for earning profits. [Figure 24.7](#) (b) shows the aggregate supply curve shifting to the left, from $SRAS_0$ to $SRAS_1$, causing the equilibrium to move from E_0 to E_1 . The movement from the original equilibrium of E_0 to the new equilibrium of E_1 will bring a nasty set of effects: reduced GDP or recession, higher unemployment because the economy is

now further away from potential GDP, and an inflationary higher price level as well. For example, the U.S. economy experienced recessions in 1974–1975, 1980–1982, 1990–91, 2001, and 2007–2009 that were each preceded or accompanied by a rise in the key input of oil prices. In the 1970s, this pattern of a shift to the left in SRAS leading to a stagnant economy with high unemployment and inflation was nicknamed **stagflation**.

Conversely, a decline in the price of a key input like oil will shift the SRAS curve to the right, providing an incentive for more to be produced at every given price level for outputs. From 1985 to 1986, for example, the average price of crude oil fell by almost half, from \$24 a barrel to \$12 a barrel. Similarly, from 1997 to 1998, the price of a barrel of crude oil dropped from \$17 per barrel to \$11 per barrel. In both cases, the plummeting oil price led to a situation like that which we presented earlier in [Figure 24.7](#) (a), where the outward shift of SRAS to the right allowed the economy to expand, unemployment to fall, and inflation to decline.

Along with energy prices, two other key inputs that may shift the SRAS curve are the cost of labor, or wages, and the cost of imported goods that we use as inputs for other products. In these cases as well, the lesson is that lower prices for inputs cause SRAS to shift to the right, while higher prices cause it to shift back to the left. Note that, unlike changes in productivity, changes in input prices do not generally cause LRAS to shift, only SRAS.

Other Supply Shocks

The aggregate supply curve can also shift due to shocks to input goods or labor. For example, an unexpected early freeze could destroy a large number of agricultural crops, a shock that would shift the AS curve to the left since there would be fewer agricultural products available at any given price.

Similarly, shocks to the labor market can affect aggregate supply. An extreme example might be an overseas war that required a large number of workers to cease their ordinary production in order to go fight for their country. In this case, SRAS and LRAS would both shift to the left because there would be fewer workers available to produce goods at any given price.

Another example in this vein is a pandemic, like the COVID-19 pandemic. A pandemic causes many workers to become sick, temporarily reducing the supply of workers by a large amount. Further, workers might be cautious to go back to work in a pandemic because of health or safety concerns. While the shock to labor supply might not be permanent, it can cause a reduction in the supply of many goods and services, reflected in a leftward shift in the short-run aggregate supply curve. At various points during the COVID-19-induced pandemic, computer chips for automobiles, meat, and other consumer services were in short supply because of worker shortages around the world.

24.4 Shifts in Aggregate Demand

LEARNING OBJECTIVES

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

- Explain how imports influence aggregate demand
- Identify ways in which business confidence and consumer confidence can affect aggregate demand
- Explain how government policy can change aggregate demand
- Evaluate why economists disagree on the topic of tax cuts

As we mentioned previously, the components of aggregate demand are consumption spending (C), investment spending (I), government spending (G), and spending on exports (X) minus imports (M). (Read the following Clear It Up feature for explanation of why imports are subtracted from exports and what this means for aggregate demand.) A shift of the AD curve to the right means that at least one of these components increased so that a greater amount of total spending would occur at every price level. A shift of the AD curve to the left means that at least one of these components decreased so that a lesser amount of total spending would occur at every price level. [The Keynesian Perspective](#) will discuss the components of aggregate demand and the factors that affect them. Here, the discussion will sketch two broad categories that could cause AD curves to

shift: changes in consumer or firm behavior and changes in government tax or spending policy.



CLEAR IT UP

Do imports diminish aggregate demand?

We have seen that the formula for aggregate demand is $AD = C + I + G + X - M$, where M is the total value of imported goods. Why is there a minus sign in front of imports? Does this mean that more imports will result in a lower level of aggregate demand? The short answer is yes, because aggregate demand is defined as total demand for domestically produced goods and services.

When an American buys a foreign product, for example, it gets counted along with all the other consumption. Thus, the income generated does not go to American producers, but rather to producers in another country. It would be wrong to count this as part of domestic demand. Therefore, imports added in consumption are subtracted back out in the M term of the equation.

Because of the way in which we write the demand equation, it is easy to make the mistake of thinking that imports are bad for the economy. Just keep in mind that every negative number in the M term has a corresponding positive number in the C or I or G term, and they always cancel out.

How Changes by Consumers and Firms Can Affect AD

When consumers feel more confident about the future of the economy, they tend to consume more. If business confidence is high, then firms tend to spend more on investment, believing that the future payoff from that investment will be substantial. Conversely, if consumer or business confidence drops, then consumption and investment spending decline.

The University of Michigan publishes a survey of consumer confidence and constructs an index of consumer confidence each month. The survey results are then reported at <http://www.sca.isr.umich.edu> (<http://www.sca.isr.umich.edu/>), which break down the change in consumer confidence among different income levels. According to that index, consumer confidence averaged around 90 prior to the Great Recession, and then it fell to below 60 in late 2008, which was the lowest it had been since 1980. During the 2010s, confidence has climbed from a 2011 low of 55.8 back to a level in the upper 90s, before falling to the lower 70s in 2020 due to the COVID-19 pandemic, which economists consider close to a healthy state.

The Organization for Economic Development and Cooperation (OECD) publishes one measure of business confidence: the "business tendency surveys". The OECD collects business opinion survey data for 21 countries on future selling prices and employment, among other business climate elements. After sharply declining during the Great Recession, the measure has risen above zero again and is back to long-term averages (the indicator dips below zero when business outlook is weaker than usual). Of course, either of these survey measures is not very precise. They can however, suggest when confidence is rising or falling, as well as when it is relatively high or low compared to the past.

Because economists associate a rise in confidence with higher consumption and investment demand, it will lead to an outward shift in the AD curve, and a move of the equilibrium, from E_0 to E_1 , to a higher quantity of output and a higher price level, as [Figure 24.8](#) (a) shows.

Consumer and business confidence often reflect macroeconomic realities; for example, confidence is usually high when the economy is growing briskly and low during a recession. However, economic confidence can sometimes rise or fall for reasons that do not have a close connection to the immediate economy, like a risk of war, election results, foreign policy events, or a pessimistic prediction about the future by a prominent public figure. U.S. presidents, for example, must be careful in their public pronouncements about the economy. If they offer economic pessimism, they risk provoking a decline in confidence that reduces consumption and investment and shifts AD to the left, and in a self-fulfilling prophecy, contributes to causing the recession that

the president warned against in the first place. Figure 24.8 (b) shows a shift of AD to the left, and the corresponding movement of the equilibrium, from E_0 to E_1 , to a lower quantity of output and a lower price level.

LINK IT UP

Visit this [website \(http://openstax.org/l/consumerconfid\)](http://openstax.org/l/consumerconfid) for data on consumer confidence.

LINK IT UP

Visit this [website \(http://openstax.org/l/businessconfid\)](http://openstax.org/l/businessconfid) for data on business confidence.

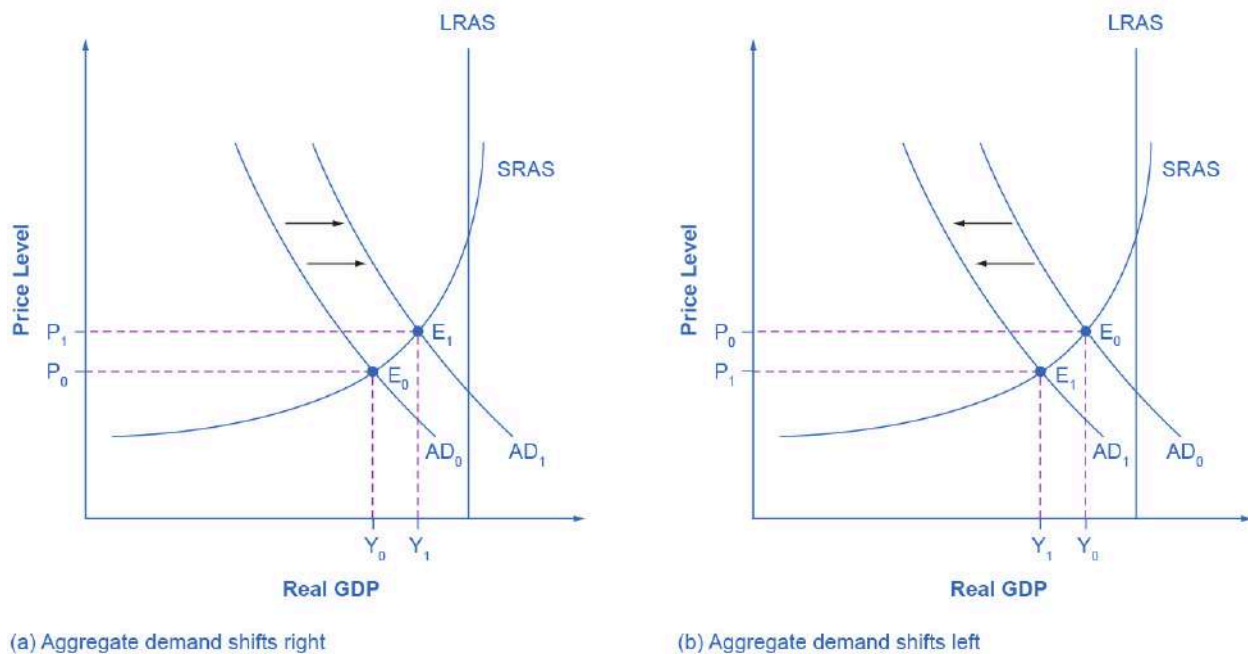


FIGURE 24.8 Shifts in Aggregate Demand (a) An increase in consumer confidence or business confidence can shift AD to the right, from AD_0 to AD_1 . When AD shifts to the right, the new equilibrium (E_1) will have a higher quantity of output and also a higher price level compared with the original equilibrium (E_0). In this example, the new equilibrium (E_1) is also closer to potential GDP. An increase in government spending or a cut in taxes that leads to a rise in consumer spending can also shift AD to the right. (b) A decrease in consumer confidence or business confidence can shift AD to the left, from AD_0 to AD_1 . When AD shifts to the left, the new equilibrium (E_1) will have a lower quantity of output and also a lower price level compared with the original equilibrium (E_0). In this example, the new equilibrium (E_1) is also farther below potential GDP. A decrease in government spending or higher taxes that leads to a fall in consumer spending can also shift AD to the left.

How Government Macroeconomic Policy Choices Can Shift AD

Government spending is one component of AD. Thus, higher government spending will cause AD to shift to the right, as in Figure 24.8 (a), while lower government spending will cause AD to shift to the left, as in Figure 24.8 (b). For example, in the United States, government spending declined by 3.2% of GDP during the 1990s, from 21% of GDP in 1991, and to 17.8% of GDP in 1998. However, from 2005 to 2009, the peak of the Great Recession, government spending increased from 19% of GDP to 21.4% of GDP. If changes of a few percentage points of GDP seem small to you, remember that since GDP was about \$14.4 trillion in 2009, a seemingly small change of 2% of GDP is equal to close to \$300 billion. Since 2009, government expenditures have gone back down to around 17–18% of GDP, although in 2020 they rose to 18.5%.

Tax policy can affect consumption and investment spending, too. Tax cuts for individuals will tend to increase consumption demand, while tax increases will tend to diminish it. Tax policy can also pump up investment demand by offering lower tax rates for corporations or tax reductions that benefit specific kinds of investment. Shifting C or I will shift the AD curve as a whole.

During a recession, when unemployment is high and many businesses are suffering low profits or even losses, the U.S. Congress often passes tax cuts. During the 2001 recession, for example, the U.S. Congress enacted a tax cut into law. At such times, the political rhetoric often focuses on how people experiencing hard times need relief from taxes. The aggregate supply and aggregate demand framework, however, offers a complementary rationale, as [Figure 24.9](#) illustrates. The original equilibrium during a recession is at point E_0 , relatively far from the full employment level of output. The tax cut, by increasing consumption, shifts the AD curve to the right. At the new equilibrium (E_1), real GDP rises and unemployment falls and, because in this diagram the economy has not yet reached its potential or full employment level of GDP, any rise in the price level remains muted. Read the following Clear It Up feature to consider the question of whether economists favor tax cuts or oppose them.

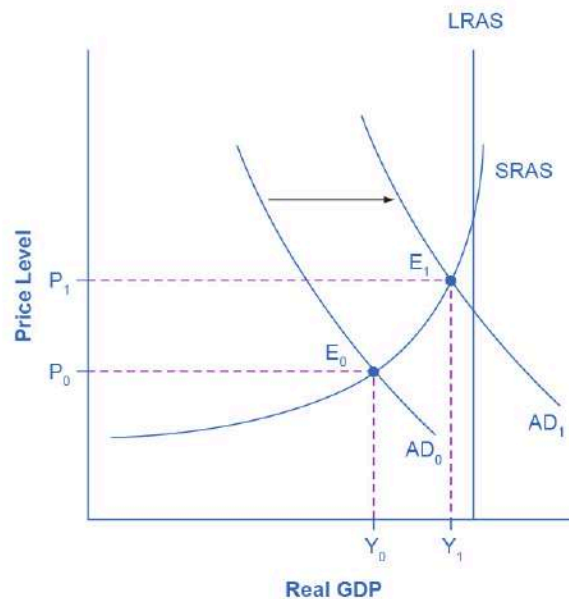


FIGURE 24.9 Recession and Full Employment in the AD/AS Model Whether the economy is in a recession is illustrated in the AD/AS model by how close the equilibrium is to the potential GDP line as indicated by the vertical LRAS line. In this example, the level of output Y_0 at the equilibrium E_0 is relatively far from the potential GDP line, so it can represent an economy in recession, well below the full employment level of GDP. In contrast, the level of output Y_1 at the equilibrium E_1 is relatively close to potential GDP, and so it would represent an economy with a lower unemployment rate.



CLEAR IT UP

Do economists favor tax cuts or oppose them?

One of the most fundamental divisions in American politics over the last few decades has been between those who believe that the government should cut taxes substantially and those who disagree. Ronald Reagan rode into the presidency in 1980 partly because of his promise, soon carried out, to enact a substantial tax cut. George Bush lost his bid for reelection against Bill Clinton in 1992 partly because he had broken his 1988 promise: “Read my lips! No new taxes!” In the 2000 presidential election, both George W. Bush and Al Gore advocated substantial tax cuts and Bush succeeded in pushing a tax cut package through Congress early in 2001. More recently in 2017 and 2018, Donald Trump initiated a new round of tax cuts throughout the economy, and President Biden promised his own set

of tax cuts in his 2021 spending bills.

What side do economists take? Do they support broad tax cuts or oppose them? The answer, unsatisfying to zealots on both sides, is that it depends. One issue is whether equally large government spending cuts accompany the tax cuts. Economists differ, as does any broad cross-section of the public, on how large government spending should be and what programs the government might cut back. A second issue, more relevant to the discussion in this chapter, concerns how close the economy is to the full employment output level. In a recession, when the AD and AS curves intersect far below the full employment level, tax cuts can make sense as a way of shifting AD to the right. However, when the economy is already performing extremely well, tax cuts may shift AD so far to the right as to generate inflationary pressures, with little gain to GDP.

With the AD/AS framework in mind, many economists might readily believe that the 1981 Reagan tax cuts, which took effect just after two serious recessions, were beneficial economic policy. Similarly, Congress enacted the 2001 Bush tax cuts and the 2009 Obama tax cuts during recessions. However, some of the same economists who favor tax cuts during recession would be much more dubious about identical tax cuts at a time the economy is performing well and cyclical unemployment is low.

Government spending and tax rate changes can be useful tools to affect aggregate demand. We will discuss these in greater detail in the [Government Budgets and Fiscal Policy](#) chapter and [The Impacts of Government Borrowing](#). Other policy tools can shift the aggregate demand curve as well. For example, as we will discuss in the [Monetary Policy and Bank Regulation](#) chapter, the Federal Reserve can affect interest rates and credit availability. Higher interest rates tend to discourage borrowing and thus reduce both household spending on big-ticket items like houses and cars and investment spending by business. Conversely, lower interest rates will stimulate consumption and investment demand. Interest rates can also affect exchange rates, which in turn will have effects on the export and import components of aggregate demand.

Clarifying the details of these alternative policies and how they affect the components of aggregate demand can wait for [The Keynesian Perspective](#) chapter. Here, the key lesson is that a shift of the aggregate demand curve to the right leads to a greater real GDP and to upward pressure on the price level. Conversely, a shift of aggregate demand to the left leads to a lower real GDP and a lower price level. Whether these changes in output and price level are relatively large or relatively small, and how the change in equilibrium relates to potential GDP, depends on whether the shift in the AD curve is happening in the AS curve's relatively flat or relatively steep portion.

24.5 How the AD/AS Model Incorporates Growth, Unemployment, and Inflation

LEARNING OBJECTIVES

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

- Use the aggregate demand/aggregate supply model to show periods of economic growth and recession
- Explain how unemployment and inflation impact the aggregate demand/aggregate supply model
- Evaluate the importance of the aggregate demand/aggregate supply model

The AD/AS model can convey a number of interlocking relationships between the three macroeconomic goals of growth, unemployment, and low inflation. Moreover, the AD/AS framework is flexible enough to accommodate both the Keynes' law approach that focuses on aggregate demand and the short run, while also including the Say's law approach that focuses on aggregate supply and the long run. These advantages are considerable. Every model is a simplified version of the deeper reality and, in the context of the AD/AS model, the three macroeconomic goals arise in ways that are sometimes indirect or incomplete. In this module, we consider how the AD/AS model illustrates the three macroeconomic goals of economic growth, low unemployment, and low inflation.

Growth and Recession in the AD/AS Diagram

In the AD/AS diagram, long-run economic growth due to productivity increases over time will be represented by a gradual shift to the right of aggregate supply. The vertical line representing potential GDP (or the “full employment level of GDP”) will gradually shift to the right over time as well. Earlier [Figure 24.7](#) (a) showed a pattern of economic growth over three years, with the AS curve shifting slightly out to the right each year. However, the factors that determine the speed of this long-term economic growth rate—like investment in physical and human capital, technology, and whether an economy can take advantage of catch-up growth—do not appear directly in the AD/AS diagram.

In the short run, GDP falls and rises in every economy, as the economy dips into recession or expands out of recession. The AD/AS diagram illustrates recessions when the equilibrium level of real GDP is substantially below potential GDP, as we see at the equilibrium point E_0 in [Figure 24.9](#). From another standpoint, in years of resurgent economic growth the equilibrium will typically be close to potential GDP, as equilibrium point E_1 in that earlier figure shows.

Unemployment in the AD/AS Diagram

We described two types of unemployment in the [Unemployment](#) chapter. Short run variations in unemployment (cyclical unemployment) are caused by the business cycle as the economy expands and contracts. Over the long run, in the United States, the unemployment rate typically hovers around 5% (give or take one percentage point or so), when the economy is healthy. In many of the national economies across Europe, the unemployment rate in recent decades has only dropped to about 10% or a bit lower, even in good economic years. We call this baseline level of unemployment that occurs year-in and year-out the natural rate of unemployment and we determine it by how well the structures of market and government institutions in the economy lead to a matching of workers and employers in the labor market. Potential GDP can imply different unemployment rates in different economies, depending on the natural rate of unemployment for that economy.

The AD/AS diagram shows cyclical unemployment by how close the economy is to the potential or full GDP employment level. Returning to [Figure 24.9](#), relatively low cyclical unemployment for an economy occurs when the level of output is close to potential GDP, as in the equilibrium point E_1 . Conversely, high cyclical unemployment arises when the output is substantially to the left of potential GDP on the AD/AS diagram, as at the equilibrium point E_0 . Although we do not show the factors that determine the natural rate of unemployment separately in the AD/AS model, they are implicitly part of what determines potential GDP or full employment GDP in a given economy.

Inflationary Pressures in the AD/AS Diagram

Inflation fluctuates in the short run. Higher inflation rates have typically occurred either during or just after economic booms: for example, the biggest spurts of inflation in the U.S. economy during the twentieth century followed the wartime booms of World War I and World War II. Conversely, rates of inflation generally decline during recessions. As an extreme example, inflation actually became negative—a situation called “deflation”—during the Great Depression. Even during the relatively short 1991–1992 recession, the inflation rate declined from 5.4% in 1990 to 3.0% in 1992. During the relatively short 2001 recession, the rate of inflation declined from 3.4% in 2000 to 1.6% in 2002. During the deep recession of 2007–2009, the inflation rate declined from 3.8% in 2008 to –0.4% in 2009. Some countries have experienced bouts of high inflation that lasted for years. In the U.S. economy since the mid-1980s, inflation does not seem to have had any long-term trend to be substantially higher. Instead, it has stayed in the 1–5% range annually.

The AD/AS framework implies two ways that inflationary pressures may arise. One possible trigger is if aggregate demand continues to shift to the right when the economy is already at or near potential GDP and full employment, thus pushing the macroeconomic equilibrium into the AS curve's steep portion. In [Figure 24.10](#) (a), there is a shift of aggregate demand to the right. The new equilibrium E_1 is clearly at a higher price level

than the original equilibrium E_0 . In this situation, the aggregate demand in the economy has soared so high that firms in the economy are not capable of producing additional goods, because labor and physical capital are fully employed, and so additional increases in aggregate demand can only result in a rise in the price level.

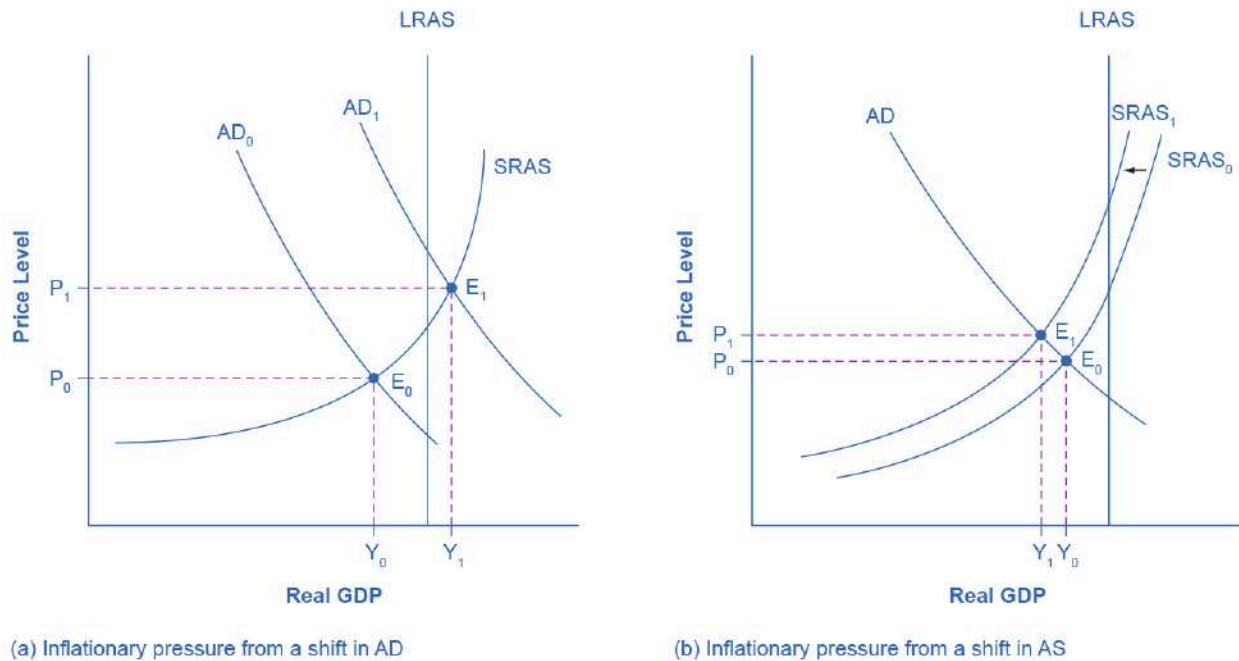


FIGURE 24.10 Sources of Inflationary Pressure in the AD/AS Model (a) A shift in aggregate demand, from AD_0 to AD_1 , when it happens in the area of the SRAS curve that is near potential GDP, will lead to a higher price level and to pressure for a higher price level and inflation. The new equilibrium (E_1) is at a higher price level (P_1) than the original equilibrium. (b) A shift in aggregate supply, from $SRAS_0$ to $SRAS_1$, will lead to a lower real GDP and to pressure for a higher price level and inflation. The new equilibrium (E_1) is at a higher price level (P_1), while the original equilibrium (E_0) is at the lower price level (P_0).

An alternative source of inflationary pressures can occur due to a rise in input prices that affects many or most firms across the economy—perhaps an important input to production like oil or labor—and causes the aggregate supply curve to shift back to the left. In [Figure 24.10](#) (b), the SRAS curve's shift to the left also increases the price level from P_0 at the original equilibrium (E_0) to a higher price level of P_1 at the new equilibrium (E_1). In effect, the rise in input prices ends up, after the final output is produced and sold, passing along in the form of a higher price level for outputs.

The AD/AS diagram shows only a one-time shift in the price level. It does not address the question of what would cause inflation either to vanish after a year, or to sustain itself for several years. There are two explanations for why inflation may persist over time. One way that continual inflationary price increases can occur is if the government continually attempts to stimulate aggregate demand in a way that keeps pushing the AD curve when it is already in the SRAS curve's steep portion. A second possibility is that, if inflation has been occurring for several years, people might begin to expect a certain level of inflation. If they do, then these expectations will cause prices, wages and interest rates to increase annually by the amount of the inflation expected. These two reasons are interrelated, because if a government fosters a macroeconomic environment with inflationary pressures, then people will grow to expect inflation. However, the AD/AS diagram does not show these patterns of ongoing or expected inflation in a direct way.

Importance of the Aggregate Demand/Aggregate Supply Model

Macroeconomics takes an overall view of the economy, which means that it needs to juggle many different concepts. For example, start with the three macroeconomic goals of growth, low inflation, and low

unemployment. Aggregate demand has four elements: consumption, investment, government spending, and exports less imports. Aggregate supply reveals how businesses throughout the economy will react to a higher price level for outputs. Finally, a wide array of economic events and policy decisions can affect aggregate demand and aggregate supply, including government tax and spending decisions; consumer and business confidence; changes in prices of key inputs like oil; and technology that brings higher levels of productivity.

The aggregate demand/aggregate supply model is one of the fundamental diagrams in this course (like the budget constraint diagram that we introduced in the [Choice in a World of Scarcity](#) chapter and the supply and demand diagram in the [Demand and Supply](#) chapter) because it provides an overall framework for bringing these factors together in one diagram. Some version of the AD/AS model will appear in every chapter in the rest of this book.

24.6 Keynes' Law and Say's Law in the AD/AS Model

LEARNING OBJECTIVES

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

- Identify the neoclassical zone, the intermediate zone, and the Keynesian zone in the aggregate demand/aggregate supply model
- Use an aggregate demand/aggregate supply model as a diagnostic test to understand the current state of the economy

We can use the AD/AS model to illustrate both Say's law that supply creates its own demand and Keynes' law that demand creates its own supply. Consider the SRAS curve's three zones which [Figure 24.11](#) identifies: the Keynesian zone, the neoclassical zone, and the intermediate zone.

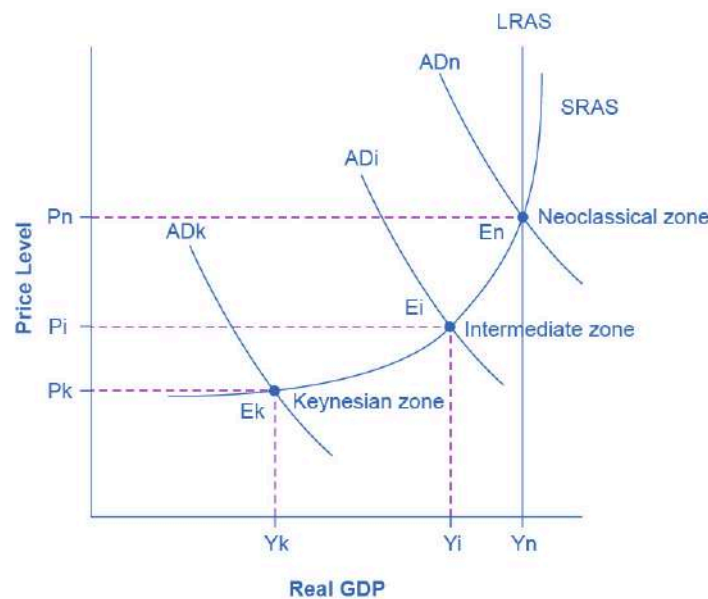


FIGURE 24.11 Keynes, Neoclassical, and Intermediate Zones in the Aggregate Supply Curve Near the equilibrium E_k , in the Keynesian zone at the far left of the SRAS curve, small shifts in AD, either to the right or the left, will affect the output level Y_k , but will not much affect the price level. In the Keynesian zone, AD largely determines the quantity of output. Near the equilibrium E_n , in the neoclassical zone at the SRAS curve's far right, small shifts in AD, either to the right or the left, will have relatively little effect on the output level Y_n , but instead will have a greater effect on the price level. In the neoclassical zone, the near-vertical SRAS curve close to the level of potential GDP largely determines the quantity of output. In the intermediate zone around equilibrium E_i , movement in AD to the right will increase both the output level and the price level, while a movement in AD to the left would decrease both the output level and the price level.

Focus first on the **Keynesian zone**, that portion of the SRAS curve on the far left which is relatively flat. If the AD curve crosses this portion of the SRAS curve at an equilibrium point like E_k , then certain statements about the economic situation will follow. In the Keynesian zone, the equilibrium level of real GDP is far below potential GDP, the economy is in recession, and cyclical unemployment is high. If aggregate demand shifted to the right or left in the Keynesian zone, it will determine the resulting level of output (and thus unemployment). However, inflationary price pressure is not much of a worry in the Keynesian zone, since the price level does not vary much in this zone.

Now, focus your attention on the **neoclassical zone** of the SRAS curve, which is the near-vertical portion on the right-hand side. If the AD curve crosses this portion of the SRAS curve at an equilibrium point like E_n where output is at or near potential GDP, then the size of potential GDP pretty much determines the level of output in the economy. Since the equilibrium is near potential GDP, cyclical unemployment is low in this economy, although structural unemployment may remain an issue. In the neoclassical zone, shifts of aggregate demand to the right or the left have little effect on the level of output or employment. The only way to increase the size of the real GDP in the neoclassical zone is for AS to shift to the right. However, shifts in AD in the neoclassical zone will create pressures to change the price level.

Finally, consider the SRAS curve's **intermediate zone** in [Figure 24.11](#). If the AD curve crosses this portion of the SRAS curve at an equilibrium point like E_i , then we might expect unemployment and inflation to move in opposing directions. For instance, a shift of AD to the right will move output closer to potential GDP and thus reduce unemployment, but will also lead to a higher price level and upward pressure on inflation. Conversely, a shift of AD to the left will move output further from potential GDP and raise unemployment, but will also lead to a lower price level and downward pressure on inflation.

This approach of dividing the SRAS curve into different zones works as a diagnostic test that we can apply to an economy, like a doctor checking a patient for symptoms. First, figure out in what zone the economy is. This will clarify the economic issues, tradeoffs, and policy choices. Some economists believe that the economy is strongly predisposed to be in one zone or another. Thus, hard-line Keynesian economists believe that the economies are in the Keynesian zone most of the time, and so they view the neoclassical zone as a theoretical abstraction. Conversely, hard-line neoclassical economists argue that economies are in the neoclassical zone most of the time and that the Keynesian zone is a distraction. [The Keynesian Perspective](#) and [The Neoclassical Perspective](#) should help to clarify the underpinnings and consequences of these contrasting views of the macroeconomy.



BRING IT HOME

The Pandemic-Induced Recession: Supply or Demand?

We mentioned earlier that a pandemic could cause a shock in the short- or long-run aggregate supply curve by temporarily reducing labor supply and slowing or stopping production of goods and services. Pandemics can also affect aggregate demand. When people are hesitant to spend or travel, or if they are not allowed to spend or travel because of social restrictions, this will affect spending in the economy. Consumers spend less at restaurants, hotels, and travel, among other areas, while firms stop investing because of the lack of demand and an uncertain future. Both actions lead to a leftward shift in the aggregate demand curve.

While there is some debate over whether the pandemic-induced recession that the U.S. economy experienced in 2020 was primarily a supply- or demand-driven one, most likely, it is a combination of both. In March and April 2020, workers left the labor market en masse, and later in the year, they were hesitant to return due to health and safety concerns. Many people were also forced to cancel travel plans or voluntarily did so out of concern for their safety, further reducing aggregate demand. These changes caused deep cuts in the global economy that continued to be felt two years after the initial pandemic-induced shocks.

Key Terms

aggregate demand (AD) the amount of total spending on domestic goods and services in an economy

aggregate demand (AD) curve the total spending on domestic goods and services at each price level

aggregate demand/aggregate supply model a model that shows what determines total supply or total demand for the economy, and how total demand and total supply interact at the macroeconomic level

aggregate supply (AS) the total quantity of output (i.e. real GDP) firms will produce and sell

aggregate supply (AS) curve the total quantity of output (i.e. real GDP) that firms will produce and sell at each price level

full-employment GDP another name for potential GDP, when the economy is producing at its potential and unemployment is at the natural rate of unemployment

intermediate zone portion of the SRAS curve where GDP is below potential but not so far below as in the Keynesian zone; the SRAS curve is upward-sloping, but not vertical in the intermediate zone

Keynes' law "demand creates its own supply"

Keynesian zone portion of the SRAS curve where GDP is far below potential and the SRAS curve is flat

long run aggregate supply (LRAS) curve vertical line at potential GDP showing no relationship between the price level for output and real GDP in the long run

neoclassical economists economists who generally emphasize the importance of aggregate supply in determining the size of the macroeconomy over the long run

neoclassical zone portion of the SRAS curve where GDP is at or near potential output where the SRAS curve is steep

potential GDP the maximum quantity that an economy can produce given full employment of its existing levels of labor, physical capital, technology, and institutions

Say's law "supply creates its own demand"

short run aggregate supply (SRAS) curve positive short run relationship between the price level for output and real GDP, holding the prices of inputs fixed

stagflation an economy experiences stagnant growth and high inflation at the same time

Key Concepts and Summary

24.1 Macroeconomic Perspectives on Demand and Supply

Neoclassical economists emphasize Say's law, which holds that supply creates its own demand. Keynesian economists emphasize Keynes' law, which holds that demand creates its own supply. Many mainstream economists take a Keynesian perspective, emphasizing the importance of aggregate demand, for the short run, and a neoclassical perspective, emphasizing the importance of aggregate supply, for the long run.

24.2 Building a Model of Aggregate Demand and Aggregate Supply

The upward-sloping short run aggregate supply (SRAS) curve shows the positive relationship between the price level and the level of real GDP in the short run. Aggregate supply slopes up because when the price level for outputs increases, while the price level of inputs remains fixed, the opportunity for additional profits encourages more production. The aggregate supply curve is near-horizontal on the left and near-vertical on the right. In the long run, we show the aggregate supply by a vertical line at the level of potential output, which is the maximum level of output the economy can produce with its existing levels of workers, physical capital, technology, and economic institutions.

The downward-sloping aggregate demand (AD) curve shows the relationship between the price level for outputs and the quantity of total spending in the economy. It slopes down because of: (a) the wealth effect, which means that a higher price level leads to lower real wealth, which reduces the level of consumption; (b) the interest rate effect, which holds that a higher price level will mean a greater demand for money, which will tend to drive up interest rates and reduce investment spending; and (c) the foreign price effect, which holds that a rise in the price level will make domestic goods relatively more expensive, discouraging exports and

encouraging imports.

24.3 Shifts in Aggregate Supply

The aggregate demand/aggregate supply (AD/AS) diagram shows how AD and AS interact. The intersection of the AD and AS curves shows the equilibrium output and price level in the economy. Movements of either AS or AD will result in a different equilibrium output and price level. The aggregate supply curve will shift out to the right as productivity increases. It will shift back to the left as the price of key inputs rises, and will shift out to the right if the price of key inputs falls. If the AS curve shifts back to the left, the combination of lower output, higher unemployment, and higher inflation, called stagflation, occurs. If AS shifts out to the right, a combination of lower inflation, higher output, and lower unemployment is possible.

24.4 Shifts in Aggregate Demand

The AD curve will shift out as the components of aggregate demand—C, I, G, and X–M—rise. It will shift back to the left as these components fall. These factors can change because of different personal choices, like those resulting from consumer or business confidence, or from policy choices like changes in government spending and taxes. If the AD curve shifts to the right, then the equilibrium quantity of output and the price level will rise. If the AD curve shifts to the left, then the equilibrium quantity of output and the price level will fall. Whether equilibrium output changes relatively more than the price level or whether the price level changes relatively more than output is determined by where the AD curve intersects with the AS curve.

The AD/AS diagram superficially resembles the microeconomic supply and demand diagram on the surface, but in reality, what is on the horizontal and vertical axes and the underlying economic reasons for the shapes of the curves are very different. We can illustrate long-term economic growth in the AD/AS framework by a gradual shift of the aggregate supply curve to the right. We illustrate a recession when the intersection of AD and AS is substantially below potential GDP, while we illustrate an expanding economy when the intersection of AS and AD is near potential GDP.

24.5 How the AD/AS Model Incorporates Growth, Unemployment, and Inflation

Cyclical unemployment is relatively large in the AD/AS framework when the equilibrium is substantially below potential GDP. Cyclical unemployment is small in the AD/AS framework when the equilibrium is near potential GDP. The natural rate of unemployment, as determined by the labor market institutions of the economy, is built into what economists mean by potential GDP, but does not otherwise appear in an AD/AS diagram. The AD/AS framework shows pressures for inflation to rise or fall when the movement from one equilibrium to another causes the price level to rise or to fall. The balance of trade does not appear directly in the AD/AS diagram, but it appears indirectly in several ways. Increases in exports or declines in imports can cause shifts in AD. Changes in the price of key imported inputs to production, like oil, can cause shifts in AS. The AD/AS model is the key model we use in this book to understand macroeconomic issues.

24.6 Keynes' Law and Say's Law in the AD/AS Model

We can divide the SRAS curve into three zones. Keynes' law says demand creates its own supply, so that changes in aggregate demand cause changes in real GDP and employment. We can show Keynes' law on the horizontal Keynesian zone of the aggregate supply curve. The Keynesian zone occurs at the left of the SRAS curve where it is fairly flat, so movements in AD will affect output, but have little effect on the price level. Say's law says supply creates its own demand. Changes in aggregate demand have no effect on real GDP and employment, only on the price level. We can show Say's law on the vertical neoclassical zone of the aggregate supply curve. The neoclassical zone occurs at the right of the SRAS curve where it is fairly vertical, and so movements in AD will affect the price level, but have little impact on output. The intermediate zone in the middle of the SRAS curve is upward-sloping, so a rise in AD will cause higher output and price level, while a fall in AD will lead to a lower output and price level.

Self-Check Questions

1. Describe the mechanism by which supply creates its own demand.
2. Describe the mechanism by which demand creates its own supply.
3. The short run aggregate supply curve was constructed assuming that as the price of outputs increases, the price of inputs stays the same. How would an increase in the prices of important inputs, like energy, affect aggregate supply?
4. In the AD/AS model, what prevents the economy from achieving equilibrium at potential output?
5. Suppose the U.S. Congress passes significant immigration reform that makes it more difficult for foreigners to come to the United States to work. Use the AD/AS model to explain how this would affect the equilibrium level of GDP and the price level.
6. Suppose concerns about the size of the federal budget deficit lead the U.S. Congress to cut all funding for research and development for ten years. Assuming this has an impact on technology growth, what does the AD/AS model predict would be the likely effect on equilibrium GDP and the price level?
7. How would a dramatic increase in the value of the stock market shift the AD curve? What effect would the shift have on the equilibrium level of GDP and the price level?
8. Suppose Mexico, one of our largest trading partners and purchaser of a large quantity of our exports, goes into a recession. Use the AD/AS model to determine the likely impact on our equilibrium GDP and price level.
9. A policymaker claims that tax cuts led the economy out of a recession. Can we use the AD/AS diagram to show this?
10. Many financial analysts and economists eagerly await the press releases for the reports on the home price index and consumer confidence index. What would be the effects of a negative report on both of these? What about a positive report?
11. What impact would a decrease in the size of the labor force have on GDP and the price level according to the AD/AS model?
12. Suppose, after five years of sluggish growth, the European Union's economy picks up speed. What would be the likely impact on the U.S. trade balance, GDP, and employment?
13. Suppose the Federal Reserve begins to increase the supply of money at an increasing rate. What impact would that have on GDP, unemployment, and inflation?
14. If the economy is operating in the neoclassical zone of the SRAS curve and aggregate demand falls, what is likely to happen to real GDP?
15. If the economy is operating in the Keynesian zone of the SRAS curve and aggregate demand falls, what is likely to happen to real GDP?

Review Questions

16. What is Say's law?
17. What is Keynes' law?
18. Do neoclassical economists believe in Keynes' law or Say's law?
19. Does Say's law apply more accurately in the long run or the short run? What about Keynes' law?
20. What is on the horizontal axis of the AD/AS diagram? What is on the vertical axis?

21. What is the economic reason why the SRAS curve slopes up?
22. What are the components of the aggregate demand (AD) curve?
23. What are the economic reasons why the AD curve slopes down?
24. Briefly explain the reason for the near-horizontal shape of the SRAS curve on its far left.
25. Briefly explain the reason for the near-vertical shape of the SRAS curve on its far right.
26. What is potential GDP?
27. Name some factors that could cause the SRAS curve to shift, and say whether they would shift SRAS to the right or to the left.
28. Will the shift of SRAS to the right tend to make the equilibrium quantity and price level higher or lower? What about a shift of SRAS to the left?
29. What is stagflation?
30. Name some factors that could cause AD to shift, and say whether they would shift AD to the right or to the left.
31. Would a shift of AD to the right tend to make the equilibrium quantity and price level higher or lower? What about a shift of AD to the left?
32. How is long-term growth illustrated in an AD/AS model?
33. How is recession illustrated in an AD/AS model?
34. How is cyclical unemployment illustrated in an AD/AS model?
35. How is the natural rate of unemployment illustrated in an AD/AS model?
36. How is pressure for inflationary price increases shown in an AD/AS model?
37. What are some of the ways in which exports and imports can affect the AD/AS model?
38. What is the Keynesian zone of the SRAS curve? How much is the price level likely to change in the Keynesian zone?
39. What is the neoclassical zone of the SRAS curve? How much is the output level likely to change in the neoclassical zone?
40. What is the intermediate zone of the SRAS curve? Will a rise in output be accompanied by a rise or a fall in the price level in this zone?

Critical Thinking Questions

41. Why would an economist choose either the neoclassical perspective or the Keynesian perspective, but not both?
42. On a microeconomic demand curve, a decrease in price causes an increase in quantity demanded because the product in question is now relatively less expensive than substitute products. Explain why aggregate demand does not increase for the same reason in response to a decrease in the aggregate price level. In other words, what causes total spending to increase if it is not because goods are now cheaper?
43. Economists expect that as the labor market continues to tighten going into the latter part of 2015 that workers should begin to expect wage increases in 2015 and 2016. Assuming this occurs and it was the only development in the labor market that year, how would this affect the AS curve? What if it was also accompanied by an increase in worker productivity?

44. If new government regulations require firms to use a cleaner technology that is also less efficient than what they previously used, what would the effect be on output, the price level, and employment using the AD/AS diagram?
45. During spring 2016 the Midwestern United States, which has a large agricultural base, experiences above-average rainfall. Using the AD/AS diagram, what is the effect on output, the price level, and employment?
46. Hydraulic fracturing (fracking) has the potential to significantly increase the amount of natural gas produced in the United States. If a large percentage of factories and utility companies use natural gas, what will happen to output, the price level, and employment as fracking becomes more widely used?
47. Some politicians have suggested tying the minimum wage to the consumer price index (CPI). Using the AD/AS diagram, what effects would this policy most likely have on output, the price level, and employment?
48. If households decide to save a larger portion of their income, what effect would this have on the output, employment, and price level in the short run? What about the long run?
49. If firms become more optimistic about the future of the economy and, at the same time, innovation in 3-D printing makes most workers more productive, what is the combined effect on output, employment, and the price-level?
50. If Congress cuts taxes at the same time that businesses become more pessimistic about the economy, what is the combined effect on output, the price level, and employment using the AD/AS diagram?
51. Suppose the level of structural unemployment increases. How would you illustrate the increase in structural unemployment in the AD/AS model? *Hint:* How does structural unemployment affect potential GDP?
52. If foreign wealth-holders decide that the United States is the safest place to invest their savings, what would the effect be on the economy here? Show graphically using the AD/AS model.
53. The AD/AS model is static. It shows a snapshot of the economy at a given point in time. Both economic growth and inflation are dynamic phenomena. Suppose economic growth is 3% per year and aggregate demand is growing at the same rate. What does the AD/AS model say the inflation rate should be?
54. Explain why the short-run aggregate supply curve might be fairly flat in the Keynesian zone of the SRAS curve. How might we tell if we are in the Keynesian zone of the AS?
55. Explain why the short-run aggregate supply curve might be vertical in the neoclassical zone of the SRAS curve. How might we tell if we are in the neoclassical zone of the AS?
56. Why might it be important for policymakers to know which in zone of the SRAS curve the economy is?
57. In your view, is the economy currently operating in the Keynesian, intermediate or neoclassical portion of the economy's aggregate supply curve?
58. Are Say's law and Keynes' law necessarily mutually exclusive?

Problems

59. Review the problem in the [Work It Out](#) titled "Interpreting the AD/AS Model." Like the information provided in that feature, [Table 24.2](#) shows information on aggregate supply, aggregate demand, and the price level for the imaginary country of Xurbia.

Price Level	AD	AS
110	700	600
120	690	640
130	680	680
140	670	720
150	660	740
160	650	760
170	640	770

TABLE 24.2 Price Level:
AD/AS

- Plot the AD/AS diagram from the data. Identify the equilibrium.
- Imagine that, as a result of a government tax cut, aggregate demand becomes higher by 50 at every price level. Identify the new equilibrium.
- How will the new equilibrium alter output? How will it alter the price level? What do you think will happen to employment?

60. The imaginary country of Harris Island has the aggregate supply and aggregate demand curves as [Table 24.3](#) shows.

Price Level	AD	AS
100	700	200
120	600	325
140	500	500
160	400	570
180	300	620

TABLE 24.3 Price Level:
AD/AS

- Plot the AD/AS diagram. Identify the equilibrium.
 - Would you expect unemployment in this economy to be relatively high or low?
 - Would you expect concern about inflation in this economy to be relatively high or low?
 - Imagine that consumers begin to lose confidence about the state of the economy, and so AD becomes lower by 275 at every price level. Identify the new aggregate equilibrium.
 - How will the shift in AD affect the original output, price level, and employment?
61. [Table 24.4](#) describes Santher's economy.

Price Level	AD	AS
50	1,000	250
60	950	580
70	900	750
80	850	850
90	800	900

TABLE 24.4 Price Level:
AD/AS

- Plot the AD/AS curves and identify the equilibrium.
- Would you expect unemployment in this economy to be relatively high or low?
- Would you expect prices to be a relatively large or small concern for this economy?
- Imagine that input prices fall and so AS shifts to the right by 150 units. Identify the new equilibrium.
- How will the shift in AS affect the original output, price level, and employment?



FIGURE 25.1 Signs of a Recession Home foreclosures were just one of the many signs and symptoms of the recent Great Recession. During that time, many businesses closed and many people lost their jobs. (Credit: modification of "Foreclosure" by Taber Andrew Bain/Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Aggregate Demand in Keynesian Analysis
- The Building Blocks of Keynesian Analysis
- The Phillips Curve
- The Keynesian Perspective on Market Forces

Introduction to the Keynesian Perspective



BRING IT HOME

The Great Recession

The 2008–2009 Great Recession hit the U.S. economy hard. According to the Bureau of Labor Statistics (BLS), the number of unemployed Americans rose from 6.8 million in May 2007 to 15.4 million in October 2009. During that time, the U.S. Census Bureau estimated that approximately 170,000 small businesses closed. Mass layoffs peaked in February 2009 when employers gave 326,392 workers notice. U.S. productivity and output fell as well. Job losses, declining home values, declining incomes, and uncertainty about the future caused consumption expenditures to decrease. According to the BLS, household spending dropped by 7.8%.

Home foreclosures and the meltdown in U.S. financial markets called for immediate action by Congress, the President, and the Federal Reserve Bank. For example, the government implemented programs such as the American Restoration and Recovery Act to help millions of people by providing tax credits for homebuyers, paying

“cash for clunkers,” and extending unemployment benefits. From cutting back on spending, filing for unemployment, and losing homes, millions of people were affected by the recession. While the United States is now on the path to recovery, people will feel the impact for many years to come.

What caused this recession and what prevented the economy from spiraling further into another depression? Policymakers looked to the lessons learned from the 1930s Great Depression and to John Maynard Keynes' models to analyze the causes and find solutions to the country's economic woes. The Keynesian perspective is the subject of this chapter.

We have learned that the level of economic activity, for example output, employment, and spending, tends to grow over time. In [The Keynesian Perspective](#) we learned the reasons for this trend. [The Macroeconomic Perspective](#) pointed out that the economy tends to cycle around the long-run trend. In other words, the economy does not always grow at its average growth rate. Sometimes economic activity grows at the trend rate, sometimes it grows more than the trend, sometimes it grows less than the trend, and sometimes it actually declines. You can see this cyclical behavior in [Figure 25.2](#).

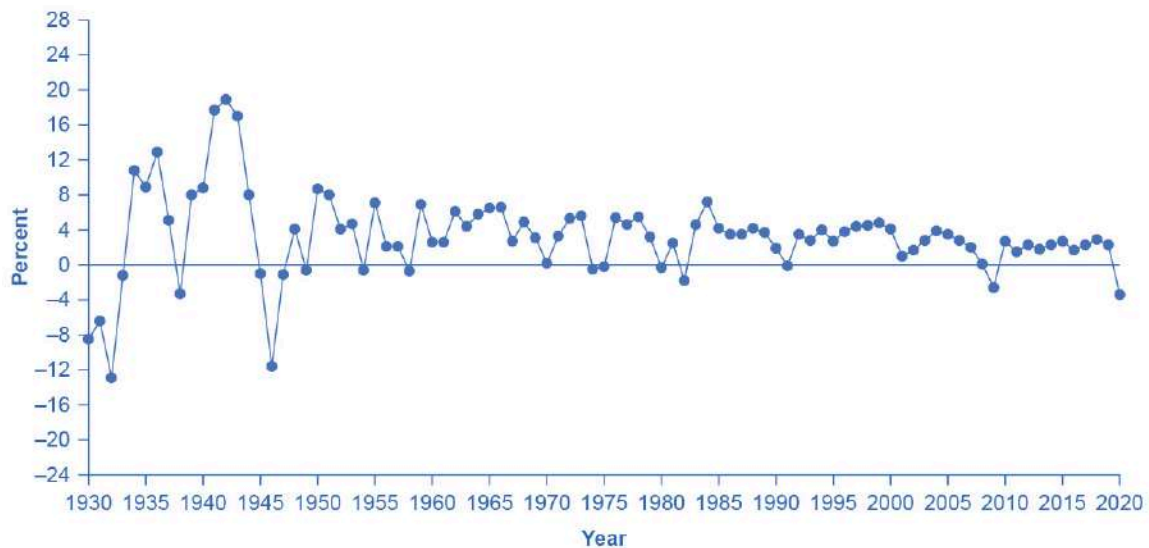


FIGURE 25.2 U.S. Real Domestic Product, Percent Changes 1930–2020 The chart tracks the percent change in Real GDP since 1930. The magnitude of both recessions and peaks was quite large between 1930 and 1945. (Source: Bureau of Economic Analysis, “National Economic Accounts,” <https://apps.bea.gov/itable/index.cfm>)

This empirical reality raises two important questions: How can we explain the cycles, and to what extent can we moderate them? This chapter (on the Keynesian perspective) and [The Neoclassical Perspective](#) explore those questions from two different points of view, building on what we learned in [The Aggregate Demand/Aggregate Supply Model](#).

[Click to view content \(https://openstax.org/books/principles-economics-3e/pages/25-introduction-to-the-keynesian-perspective\)](https://openstax.org/books/principles-economics-3e/pages/25-introduction-to-the-keynesian-perspective)

Percent Change of U.S. Real Gross Domestic Product.

25.1 Aggregate Demand in Keynesian Analysis

LEARNING OBJECTIVES

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

- Explain real GDP, recessionary gaps, and inflationary gaps
- Recognize the Keynesian AD/AS model
- Identify the determining factors of both consumption expenditure and investment expenditure
- Analyze the factors that determine government spending and net exports

The Keynesian perspective focuses on aggregate demand. The idea is simple: firms produce output only if they expect it to sell. Thus, while the availability of the factors of production determines a nation's potential GDP, the amount of goods and services that actually sell, known as **real GDP**, depends on how much demand exists across the economy. [Figure 25.3](#) illustrates this point.

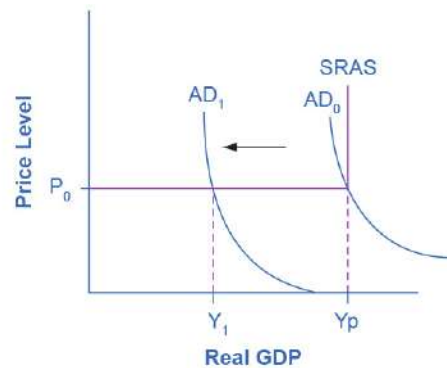


FIGURE 25.3 The Keynesian AD/AS Model The Keynesian View of the AD/AS Model uses an SRAS curve, which is horizontal at levels of output below potential and vertical at potential output. Thus, when beginning from potential output, any decrease in AD affects only output, but not prices. Any increase in AD affects only prices, not output.

Keynes argued that, for reasons we explain shortly, aggregate demand is not stable—that it can change unexpectedly. Suppose the economy starts where AD intersects SRAS at P_0 and Y_p . Because Y_p is potential output, the economy is at full employment. Because AD is volatile, it can easily fall. Thus, even if we start at Y_p , if AD falls, then we find ourselves in what Keynes termed a **recessionary gap**. The economy is in equilibrium but with less than full employment, as Y_1 in [Figure 25.3](#) shows. Keynes believed that the economy would tend to stay in a recessionary gap, with its attendant unemployment, for a significant period of time.

In the same way (although we do not show it in the figure), if AD increases, the economy could experience an **inflationary gap**, where demand is attempting to push the economy past potential output. Consequently, the economy experiences inflation. The key policy implication for either situation is that government needs to step in and close the gap, increasing spending during recessions and decreasing spending during booms to return aggregate demand to match potential output.

Recall from [The Aggregate Supply-Aggregate Demand Model](#) that aggregate demand is total spending, economy-wide, on domestic goods and services. (Aggregate demand (AD) is actually what economists call total planned expenditure. Read the appendix on [The Expenditure-Output Model](#) for more on this.) You may also remember that aggregate demand is the sum of four components: consumption expenditure, investment expenditure, government spending, and spending on net exports (exports minus imports). In the following sections, we will examine each component through the Keynesian perspective.

What Determines Consumption Expenditure?

Consumption expenditure is spending by households and individuals on durable goods, nondurable goods, and services. Durable goods are items that last and provide value over time, such as automobiles. Nondurable goods are things like groceries—once you consume them, they are gone. Recall from [The Macroeconomic](#)

Perspective that services are intangible things consumers buy, like healthcare or entertainment.

Keynes identified three factors that affect consumption:

- **Disposable income:** For most people, the single most powerful determinant of how much they consume is how much income they have in their take-home pay, also known as **disposable income**, which is income after taxes.
- **Expected future income:** Consumer expectations about future income also are important in determining consumption. If consumers feel optimistic about the future, they are more likely to spend and increase overall aggregate demand. News of recession and troubles in the economy will make them pull back on consumption.
- **Wealth or credit:** When households experience a rise in wealth, they may be willing to consume a higher share of their income and to save less. When the U.S. stock market rose dramatically in the late 1990s, for example, U.S. savings rates declined, probably in part because people felt that their wealth had increased and there was less need to save. How do people spend beyond their income, when they perceive their wealth increasing? The answer is borrowing. On the other side, when the U.S. stock market declined about 40% from March 2008 to March 2009, people felt far greater uncertainty about their economic future, so savings rates increased while consumption declined.

Finally, Keynes noted that a variety of other factors combine to determine how much people save and spend. If household preferences about saving shift in a way that encourages consumption rather than saving, then AD will shift out to the right.

LINK IT UP

Visit this [website \(http://openstax.org/l/Diane_Rehm\)](http://openstax.org/l/Diane_Rehm) for more information about how the recession affected various groups of people.

What Determines Investment Expenditure?

We call spending on new capital goods investment expenditure. Investment falls into four categories: producer's durable equipment and software, nonresidential structures (such as factories, offices, and retail locations), changes in inventories, and residential structures (such as single-family homes, townhouses, and apartment buildings). Businesses conduct the first three types of investment, while households conduct the last.

Keynes's treatment of investment focuses on the key role of expectations about the future in influencing business decisions. When a business decides to make an investment in physical assets, like plants or equipment, or in intangible assets, like skills or a research and development project, that firm considers both the expected investment benefits (future profit expectations) and the investment costs (interest rates).

- **Expectations of future profits:** The clearest driver of investment benefits is expectations for future profits. When we expect an economy to grow, businesses perceive a growing market for their products. Their higher degree of business confidence will encourage new investment. For example, in the second half of the 1990s, U.S. investment levels surged from 18% of GDP in 1994 to 21% in 2000. However, when a recession started in 2001, U.S. investment levels quickly sank back to 18% of GDP by 2002.
- **Interest rates** also play a significant role in determining how much investment a firm will make. Just as individuals need to borrow money to purchase homes, so businesses need financing when they purchase big ticket items. The cost of investment thus includes the interest rate. Even if the firm has the funds, the interest rate measures the opportunity cost of purchasing business capital. Lower interest rates stimulate investment spending and higher interest rates reduce it.

Many factors can affect the expected profitability on investment. For example, if the energy prices decline, then investments that use energy as an input will yield higher profits. If government offers special incentives

for investment (for example, through the tax code), then investment will look more attractive; conversely, if government removes special investment incentives from the tax code, or increases other business taxes, then investment will look less attractive. As Keynes noted, business investment is the most variable of all the components of aggregate demand.

What Determines Government Spending?

The third component of aggregate demand is federal, state, and local government spending. Although we usually view the United States as a market economy, government still plays a significant role in the economy. As we discuss in [Environmental Protection and Negative Externalities](#) and [Positive Externalities and Public Goods](#), government provides important public services such as national defense, transportation infrastructure, and education.

Keynes recognized that the government budget offered a powerful tool for influencing aggregate demand. Not only could more government spending stimulate AD (or less government spending reduce it), but lowering or raising tax rates could influence consumption and investment spending. Keynes concluded that during extreme times like deep recessions, only the government had the power and resources to move aggregate demand. For example, during the 2020 pandemic-induced recession, the federal government gave money to state and local governments and to households to support the economy when many firms and governments needed to shut down or suffered a large decline in revenue and needed to lay off workers.

What Determines Net Exports?

Recall that exports are domestically produced products that sell abroad while imports are foreign produced products that consumers purchase domestically. Since we define aggregate demand as spending on domestic goods and services, export expenditures add to AD, while import expenditures subtract from AD.

Two sets of factors can cause shifts in export and import demand: changes in relative growth rates between countries and changes in relative prices between countries. What is happening in the countries' economies that would be purchasing those exports heavily affects the level of demand for a nation's exports. For example, if major importers of American-made products like Canada, Japan, and Germany have recessions, exports of U.S. products to those countries are likely to decline. Conversely, the amount of income in the domestic economy directly affects the quantity of a nation's imports: more income will bring a higher level of imports.

Relative prices of goods in domestic and international markets can also affect exports and imports. If U.S. goods are relatively cheaper compared with goods made in other places, perhaps because a group of U.S. producers has mastered certain productivity breakthroughs, then U.S. exports are likely to rise. If U.S. goods become relatively more expensive, perhaps because a change in the exchange rate between the U.S. dollar and other currencies has pushed up the price of inputs to production in the United States, then exports from U.S. producers are likely to decline.

[Table 25.1](#) summarizes the reasons we have explained for changes in aggregate demand.

Reasons for a Decrease in Aggregate Demand	Reasons for an Increase in Aggregate Demand
Consumption <ul style="list-style-type: none"> • Rise in taxes • Fall in income • Rise in interest rates • Desire to save more • Decrease in wealth • Fall in future expected income 	Consumption <ul style="list-style-type: none"> • Decrease in taxes • Increase in income • Fall in interest rates • Desire to save less • Rise in wealth • Rise in future expected income
Investment <ul style="list-style-type: none"> • Fall in expected rate of return • Rise in interest rates • Drop in business confidence 	Investment <ul style="list-style-type: none"> • Rise in expected rate of return • Drop in interest rates • Rise in business confidence
Government <ul style="list-style-type: none"> • Reduction in government spending • Increase in taxes 	Government <ul style="list-style-type: none"> • Increase in government spending • Decrease in taxes
Net Exports <ul style="list-style-type: none"> • Decrease in foreign demand • Relative price increase of U.S. goods 	Net Exports <ul style="list-style-type: none"> • Increase in foreign demand • Relative price drop of U.S. goods

TABLE 25.1 Determinants of Aggregate Demand

25.2 The Building Blocks of Keynesian Analysis

LEARNING OBJECTIVES

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

- Evaluate the Keynesian view of recessions through an understanding of sticky wages and prices and the importance of aggregate demand
- Explain the coordination argument, menu costs, and macroeconomic externality
- Analyze the impact of the expenditure multiplier

Now that we have a clear understanding of what constitutes aggregate demand, we return to the Keynesian argument using the model of aggregate demand/aggregate supply (AD/AS). (For a similar treatment using Keynes' income-expenditure model, see the appendix on [The Expenditure-Output Model](#).)

Keynesian economics focuses on explaining why recessions and depressions occur and offering a policy prescription for minimizing their effects. The Keynesian view of recession is based on two key building blocks. First, aggregate demand is not always automatically high enough to provide firms with an incentive to hire enough workers to reach full employment. Second, the macroeconomy may adjust only slowly to shifts in aggregate demand because of **sticky wages and prices**, which are wages and prices that do not respond to decreases or increases in demand. We will consider these two claims in turn, and then see how they are represented in the AD/AS model.

The first building block of the Keynesian diagnosis is that recessions occur when the level of demand for goods and services is less than what is produced when labor is fully employed. In other words, the intersection of aggregate supply and aggregate demand occurs at a level of output less than the level of GDP consistent with full employment. Suppose the stock market crashes, as in 1929, or suppose the housing market collapses, as in 2008. In either case, household wealth will decline, and consumption expenditure will follow. Suppose

businesses see that consumer spending is falling or face restrictions from a pandemic that curtail their operations. That will reduce expectations of the profitability of investment, so businesses will decrease investment expenditure.

This seemed to be the case during the Great Depression, since the physical capacity of the economy to supply goods did not alter much. No flood or earthquake or other natural disaster ruined factories in 1929 or 1930. No outbreak of disease decimated the ranks of workers. No key input price, like the price of oil, soared on world markets. The U.S. economy in 1933 had just about the same factories, workers, and state of technology as it had had four years earlier in 1929—and yet the economy had shrunk dramatically. This also seems to be what happened in 2008.

As Keynes recognized, the events of the Depression contradicted Say's law that “supply creates its own demand.” Although production capacity existed, the markets were not able to sell their products. As a result, real GDP was less than potential GDP.

LINK IT UP

Visit this [website \(http://openstax.org/l/expenditures\)](http://openstax.org/l/expenditures) for raw data used to calculate GDP.

Wage and Price Stickiness

Keynes also pointed out that although AD fluctuated, prices and wages did not immediately respond as economists often expected. Instead, prices and wages are “sticky,” making it difficult to restore the economy to full employment and potential GDP. Keynes emphasized one particular reason why wages were sticky: the **coordination argument**. This argument points out that, even if most people would be willing—at least hypothetically—to see a decline in their own wages in bad economic times as long as everyone else also experienced such a decline, a market-oriented economy has no obvious way to implement a plan of coordinated wage reductions. [Unemployment](#) proposed a number of reasons why wages might be sticky downward, most of which center on the argument that businesses avoid wage cuts because they may in one way or another depress morale and hurt the productivity of the existing workers.

Some modern economists have argued in a Keynesian spirit that, along with wages, other prices may be sticky, too. Many firms do not change their prices every day or even every month. When a firm considers changing prices, it must consider two sets of costs. First, changing prices uses company resources: managers must analyze the competition and market demand and decide the new prices, they must update sales materials, change billing records, and redo product and price labels. Second, frequent price changes may leave customers confused or angry—especially if they discover that a product now costs more than they expected. These costs of changing prices are called **menu costs**—like the costs of printing a new set of menus with different prices in a restaurant. Prices do respond to forces of supply and demand, but from a macroeconomic perspective, the process of changing all prices throughout the economy takes time.

To understand the effect of sticky wages and prices in the economy, consider [Figure 25.4](#) (a) illustrating the overall labor market, while [Figure 25.4](#) (b) illustrates a market for a specific good or service. The original equilibrium (E_0) in each market occurs at the intersection of the demand curve (D_0) and supply curve (S_0). When aggregate demand declines, the demand for labor shifts to the left (to D_1) in [Figure 25.4](#) (a) and the demand for goods shifts to the left (to D_1) in [Figure 25.4](#) (b). However, because of sticky wages and prices, the wage remains at its original level (W_0) for a period of time and the price remains at its original level (P_0).

As a result, a situation of excess supply—where the quantity supplied exceeds the quantity demanded at the existing wage or price—exists in markets for both labor and goods, and Q_1 is less than Q_0 in both [Figure 25.4](#) (a) and [Figure 25.4](#) (b). When many labor markets and many goods markets all across the economy find themselves in this position, the economy is in a recession; that is, firms cannot sell what they wish to produce at the existing market price and do not wish to hire all who are willing to work at the existing market wage. The

Clear It Up feature discusses this problem in more detail.

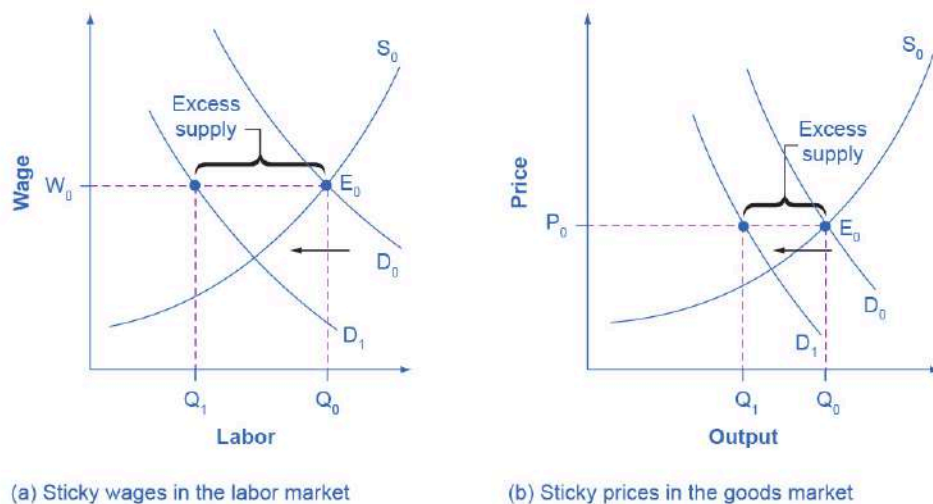


FIGURE 25.4 Sticky Prices and Falling Demand in the Labor and Goods Market In both (a) and (b), demand shifts left from D_0 to D_1 . However, the wage in (a) and the price in (b) do not immediately decline. In (a), the quantity demanded of labor at the original wage (W_0) is Q_0 , but with the new demand curve for labor (D_1), it will be Q_1 . Similarly, in (b), the quantity demanded of goods at the original price (P_0) is Q_0 , but at the new demand curve (D_1) it will be Q_1 . An excess supply of labor will exist, which we call unemployment. An excess supply of goods will also exist, where the quantity demanded is substantially less than the quantity supplied. Thus, sticky wages and sticky prices, combined with a drop in demand, bring about unemployment and recession.



CLEAR IT UP

Why Was the Pace of Wage Adjustments Slow?

The recovery after the Great Recession in the United States was slow. In fact, many low-wage workers at McDonalds, Dominos, and Walmart threatened to strike for higher wages. Their plight was part of a larger trend in job growth and pay in the post-recession recovery.

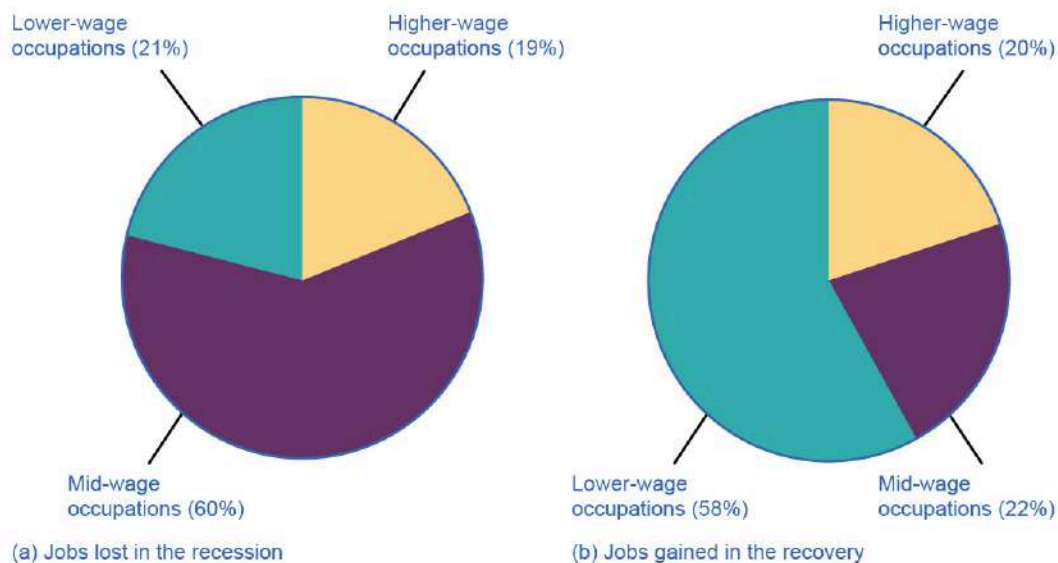


FIGURE 25.5 Jobs Lost/Gained in the Recession/Recovery Data in the aftermath of the Great Recession suggests that jobs lost were in mid-wage occupations, while jobs gained were in low-wage occupations.

The National Employment Law Project compiled data from the Bureau of Labor Statistics and found that, during the Great Recession, 60% of job losses were in medium-wage occupations. Most of them were replaced during the recovery period with lower-wage jobs in the service, retail, and food industries. [Figure 25.5](#) illustrates this data.

Wages in the service, retail, and food industries are at or near minimum wage and tend to be both downwardly and upwardly “sticky.” Wages are downwardly sticky due to minimum wage laws. They may be upwardly sticky if insufficient competition in low-skilled labor markets enables employers to avoid raising wages that would reduce their profits. At the same time, however, the Consumer Price Index increased 11% between 2007 and 2012, pushing real wages down.

The Two Keynesian Assumptions in the AD/AS Model

[Figure 25.6](#) is the AD/AS diagram which illustrates these two Keynesian assumptions—the importance of aggregate demand in causing recession and the stickiness of wages and prices. Note that because of the stickiness of wages and prices, the aggregate supply curve is flatter than either supply curve (labor or specific good). In fact, if wages and prices were so sticky that they did not fall at all, the aggregate supply curve would be completely flat below potential GDP, as [Figure 25.6](#) shows. This outcome is an important example of a **macroeconomic externality**, where what happens at the macro level is different from and inferior to what happens at the micro level. For example, a firm should respond to a decrease in demand for its product by cutting its price to increase sales. However, if all firms experience a decrease in demand for their products, sticky prices in the aggregate prevent aggregate demand from rebounding (which we would show as a movement along the AD curve in response to a lower price level).

The original equilibrium of this economy occurs where the aggregate demand function (AD_0) intersects with AS. Since this intersection occurs at potential GDP (Y_p), the economy is operating at full employment. When aggregate demand shifts to the left, all the adjustment occurs through decreased real GDP. There is no decrease in the price level. Since the equilibrium occurs at Y_1 , the economy experiences substantial unemployment.

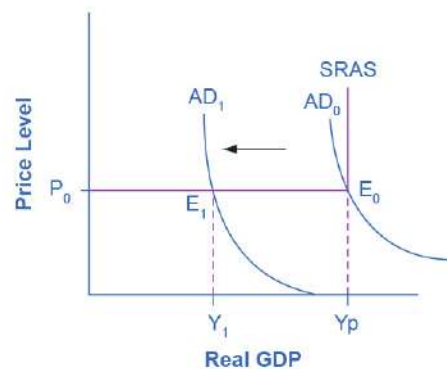


FIGURE 25.6 A Keynesian Perspective of Recession This figure illustrates the two key assumptions behind Keynesian economics. A recession begins when aggregate demand declines from AD_0 to AD_1 . The recession persists because of the assumption of fixed wages and prices, which makes the SRAS flat below potential GDP. If that were not the case, the price level would fall also, raising GDP and limiting the recession. Instead the intersection E_1 occurs in the flat portion of the SRAS curve where GDP is less than potential.

The Expenditure Multiplier

A key concept in Keynesian economics is the **expenditure multiplier**. The expenditure multiplier is the idea that not only does spending affect the equilibrium level of GDP, but that spending is powerful. More precisely, it means that a change in spending causes a more than proportionate change in GDP.

$$\frac{\Delta Y}{\Delta \text{Spending}} > 1$$

The reason for the expenditure multiplier is that one person's spending becomes another person's income, which leads to additional spending and additional income so that the cumulative impact on GDP is larger than the initial increase in spending. The appendix on [The Expenditure-Output Model](#) provides the details of the multiplier process, but the concept is important enough for us to summarize here. While the multiplier is important for understanding the effectiveness of fiscal policy, it occurs whenever any autonomous increase in spending occurs. Additionally, the multiplier operates in a negative as well as a positive direction. Thus, when investment spending collapsed during the Great Depression, it caused a much larger decrease in real GDP. The size of the multiplier is critical and was a key element in discussions of the effectiveness of the Obama administration's fiscal stimulus package, officially titled the American Recovery and Reinvestment Act of 2009.

25.3 The Phillips Curve

LEARNING OBJECTIVES

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

- Explain the Phillips curve, noting its impact on the theories of Keynesian economics
- Graph a Phillips curve
- Identify factors that cause the instability of the Phillips curve
- Analyze the Keynesian policy for reducing unemployment and inflation

The simplified AD/AS model that we have used so far is fully consistent with Keynes's original model. More recent research, though, has indicated that in the real world, an aggregate supply curve is more curved than the right angle that we used in this chapter. Rather, the real-world AS curve is very flat at levels of output far below potential ("the Keynesian zone"), very steep at levels of output above potential ("the neoclassical zone") and curved in between ("the intermediate zone"). [Figure 25.7](#) illustrates this. The typical aggregate supply curve leads to the concept of the Phillips curve.

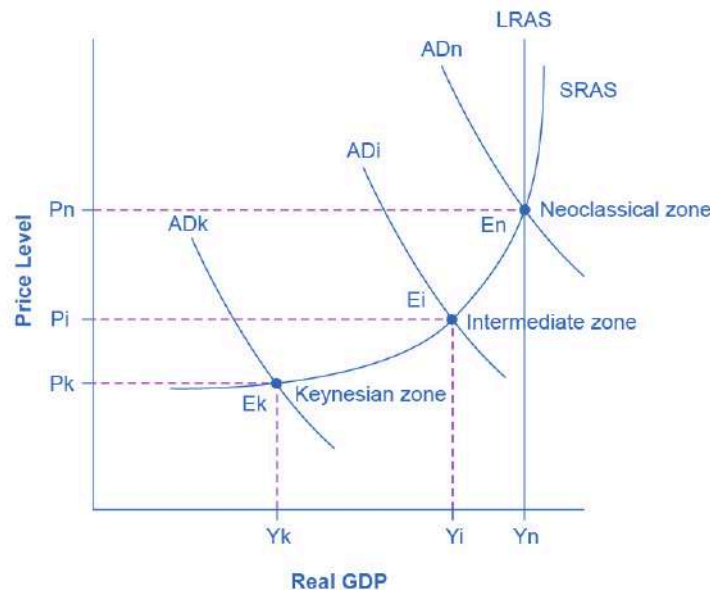


FIGURE 25.7 Keynes, Neoclassical, and Intermediate Zones in the Aggregate Supply Curve Near the equilibrium E_k , in the Keynesian zone at the SRAS curve's far left, small shifts in AD, either to the right or the left, will affect the output level Y_k , but will not much affect the price level. In the Keynesian zone, AD largely determines the quantity of output. Near the equilibrium E_n , in the neoclassical zone, at the SRAS curve's far right, small shifts in AD, either to the right or the left, will have relatively little effect on the output level Y_n , but instead will have a greater effect on the price level. In the neoclassical zone, the near-vertical SRAS curve close to the level of potential GDP (as

represented by the LRAS line) largely determines the quantity of output. In the intermediate zone around equilibrium E_i , movement in AD to the right will increase both the output level and the price level, while a movement in AD to the left would decrease both the output level and the price level.

The Discovery of the Phillips Curve

In the 1950s, A.W. Phillips, an economist at the London School of Economics, was studying the Keynesian analytical framework. The Keynesian theory implied that during a recession inflationary pressures are low, but when the level of output is at or even pushing beyond potential GDP, the economy is at greater risk for inflation. Phillips analyzed 60 years of British data and did find that tradeoff between unemployment and inflation, which became known as the **Phillips curve**. Figure 25.8 shows a theoretical Phillips curve, and the following Work It Out feature shows how the pattern appears for the United States.

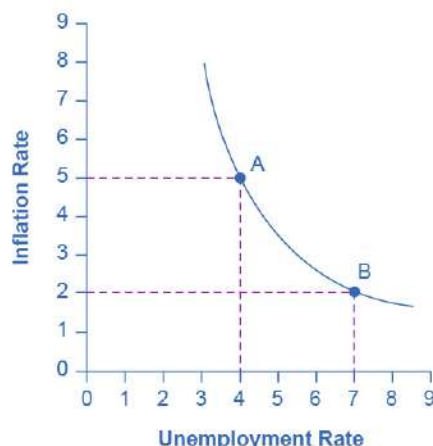


FIGURE 25.8 A Keynesian Phillips Curve Tradeoff between Unemployment and Inflation A Phillips curve illustrates a tradeoff between the unemployment rate and the inflation rate. If one is higher, the other must be lower. For example, point A illustrates a 5% inflation rate and a 4% unemployment. If the government attempts to reduce inflation to 2%, then it will experience a rise in unemployment to 7%, as point B shows.

WORK IT OUT

The Phillips Curve for the United States

Step 1. Go to this [website \(https://openstax.org/l/ERP2005\)](https://openstax.org/l/ERP2005) to see the 2005 *Economic Report of the President*.

Step 2. Scroll down and locate Table B-63 in the Appendices. This table is titled “Changes in special consumer price indexes, 1960–2004.”

Step 3. Download the table in Excel by selecting the XLS option and then selecting the location in which to save the file.

Step 4. Open the downloaded Excel file.

Step 5. View the third column (labeled “Year to year”). This is the inflation rate, measured by the percentage change in the Consumer Price Index.

Step 6. Return to the website and scroll to locate the Appendix Table B-42 “Civilian unemployment rate, 1959–2004.”

Step 7. Download the table in Excel.

Step 8. Open the downloaded Excel file and view the second column. This is the overall unemployment rate.

Step 9. Using the data available from these two tables, plot the Phillips curve for 1960–69, with unemployment rate on the x-axis and the inflation rate on the y-axis. Your graph should look like [Figure 25.9](#).

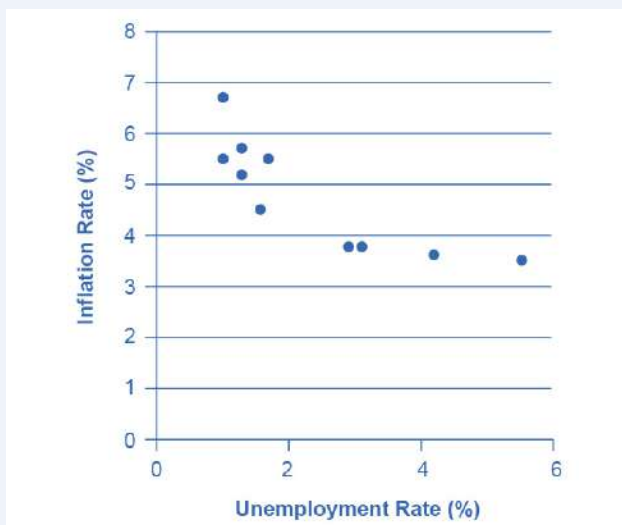


FIGURE 25.9 The Phillips Curve from 1960–1969 This chart shows the negative relationship between unemployment and inflation.

Step 10. Plot the Phillips curve for 1960–1979. What does the graph look like? Do you still see the tradeoff between inflation and unemployment? Your graph should look like [Figure 25.10](#).

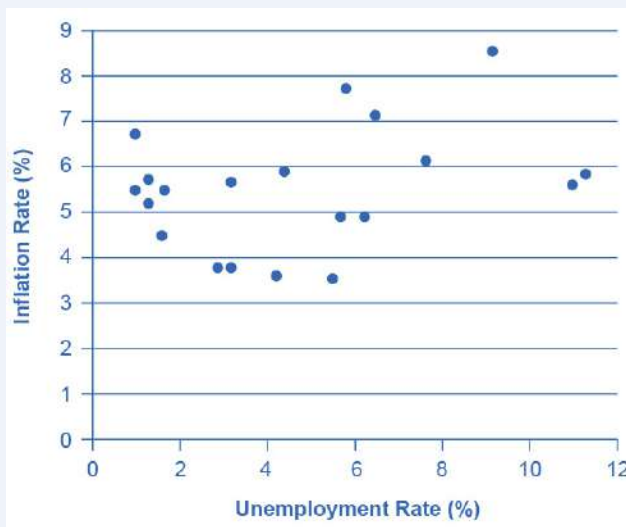


FIGURE 25.10 U.S. Phillips Curve, 1960–1979 The tradeoff between unemployment and inflation appeared to break down during the 1970s as the Phillips Curve shifted out to the right.

Over this longer period of time, the Phillips curve appears to have shifted out. There is no tradeoff any more.

The Instability of the Phillips Curve

During the 1960s, economists viewed the Phillips curve as a policy menu. A nation could choose low inflation and high unemployment, or high inflation and low unemployment, or anywhere in between. Economies could use fiscal and monetary policy to move up or down the Phillips curve as desired. Then a curious thing happened. When policymakers tried to exploit the tradeoff between inflation and unemployment, the result was an increase in both inflation and unemployment. What had happened? The Phillips curve shifted.

The U.S. economy experienced this pattern in the deep recession from 1973 to 1975, and again in back-to-back recessions from 1980 to 1982. Many nations around the world saw similar increases in unemployment and inflation. This pattern became known as stagflation. (Recall from [The Aggregate Demand/Aggregate Supply Model](#) that stagflation is an unhealthy combination of high unemployment and high inflation.) Perhaps most important, stagflation was a phenomenon that traditional Keynesian economics could not explain.

Economists have concluded that two factors cause the Phillips curve to shift. The first is supply shocks, like the mid-1970s oil crisis, which first brought stagflation into our vocabulary. The second is changes in people's expectations about inflation. In other words, there may be a tradeoff between inflation and unemployment when people expect no inflation, but when they realize inflation is occurring, the tradeoff disappears. Both factors (supply shocks and changes in inflationary expectations) cause the aggregate supply curve, and thus the Phillips curve, to shift.

In short, we should interpret a downward-sloping Phillips curve as valid for short-run periods of several years, but over longer periods, when aggregate supply shifts, the downward-sloping Phillips curve can shift so that unemployment and inflation are both higher (as in the 1970s and early 1980s) or both lower (as in the early 1990s or first decade of the 2000s).

Keynesian Policy for Fighting Unemployment and Inflation

Keynesian macroeconomics argues that the solution to a recession is **expansionary fiscal policy**, such as tax cuts to stimulate consumption and investment, or direct increases in government spending that would shift the aggregate demand curve to the right. For example, if aggregate demand was originally at AD_r in [Figure 25.11](#), so that the economy was in recession, the appropriate policy would be for government to shift aggregate demand to the right from AD_r to AD_f , where the economy would be at potential GDP and full employment.

Keynes noted that while it would be nice if the government could spend additional money on housing, roads, and other amenities, he also argued that if the government could not agree on how to spend money in practical ways, then it could spend in impractical ways. For example, Keynes suggested building monuments, like a modern equivalent of the Egyptian pyramids. He proposed that the government could bury money underground, and let mining companies start digging up the money again. These suggestions were slightly tongue-in-cheek, but their purpose was to emphasize that a Great Depression is no time to quibble over the specifics of government spending programs and tax cuts when the goal should be to pump up aggregate demand by enough to lift the economy to potential GDP.

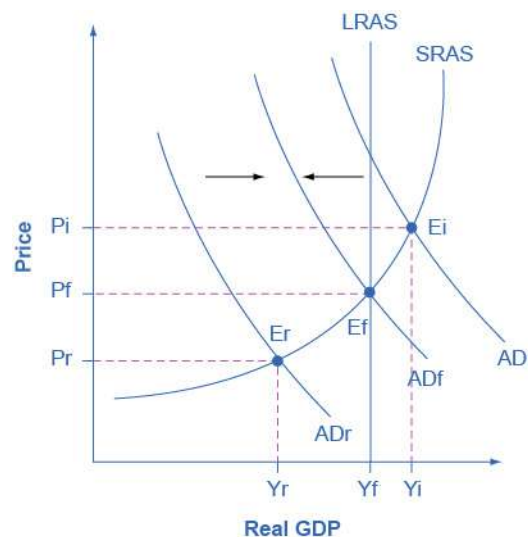


FIGURE 25.11 Fighting Recession and Inflation with Keynesian Policy If an economy is in recession, with an equilibrium at E_r , then the Keynesian response would be to enact a policy to shift aggregate demand to the right from AD_r toward AD_f . If an economy is experiencing inflationary pressures with an equilibrium at E_i , then the

Keynesian response would be to enact a policy response to shift aggregate demand to the left, from AD_i toward AD_f.

The other side of Keynesian policy occurs when the economy is operating above potential GDP. In this situation, unemployment is low, but inflationary rises in the price level are a concern. The Keynesian response would be **contractionary fiscal policy**, using tax increases or government spending cuts to shift AD to the left. The result would be downward pressure on the price level, but very little reduction in output or very little rise in unemployment. If aggregate demand was originally at AD_i in [Figure 25.11](#), so that the economy was experiencing inflationary rises in the price level, the appropriate policy would be for government to shift aggregate demand to the left, from AD_i toward AD_f, which reduces the pressure for a higher price level while the economy remains at full employment.

In the Keynesian economic model, too little aggregate demand brings unemployment and too much brings inflation. Thus, you can think of Keynesian economics as pursuing a “Goldilocks” level of aggregate demand: not too much, not too little, but looking for what is just right.

25.4 The Keynesian Perspective on Market Forces

LEARNING OBJECTIVES

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

- Explain the Keynesian perspective on market forces
- Analyze the role of government policy in economic management

Ever since the birth of Keynesian economics in the 1930s, controversy has simmered over the extent to which government should play an active role in managing the economy. In the aftermath of the human devastation and misery of the Great Depression, many people—including many economists—became more aware of vulnerabilities within the market-oriented economic system. Some supporters of Keynesian economics advocated a high degree of government planning in all parts of the economy.

However, Keynes himself was careful to separate the issue of aggregate demand from the issue of how well individual markets worked. He argued that individual markets for goods and services were appropriate and useful, but that sometimes that level of aggregate demand was just too low. When 10 million people are willing and able to work, but one million of them are unemployed, he argued, individual markets may be doing a perfectly good job of allocating the efforts of the nine million workers—the problem is that insufficient aggregate demand exists to support jobs for all 10 million. Thus, he believed that, while government should ensure that overall level of aggregate demand is sufficient for an economy to reach full employment, this task did not imply that the government should attempt to set prices and wages throughout the economy, nor to take over and manage large corporations or entire industries directly.

Even if one accepts the Keynesian economic theory, a number of practical questions remain. In the real world, can government economists identify potential GDP accurately? Is a desired increase in aggregate demand better accomplished by a tax cut or by an increase in government spending? Given the inevitable delays and uncertainties as governments enact policies into law, is it reasonable to expect that the government can implement Keynesian economics? Can fixing a recession really be just as simple as pumping up aggregate demand? [Government Budgets and Fiscal Policy](#) will probe these issues. The Keynesian approach, with its focus on aggregate demand and sticky prices, has proved useful in understanding how the economy fluctuates in the short run and why recessions and cyclical unemployment occur. In [The Neoclassical Perspective](#), we will consider some of the shortcomings of the Keynesian approach and why it is not especially well-suited for long-run macroeconomic analysis.



BRING IT HOME

The Pandemic-Induced Recession and the Keynesian Perspective

The pandemic-induced recession of 2020 was unique. Unlike the Great Recession of 2007–2009 discussed in most

of this chapter, it was not started by the burst of a housing bubble. It was started by a virus that caused sickness and death for millions of people worldwide and required substantial social policy in order to control its spread.

In some ways, the latest recession was influenced by fluctuations in aggregate demand. At its depth in April 2020, over 20 million people were unemployed, causing a massive decline in consumption and aggregate demand. As businesses were forced to shutter or move their operations online, many were pessimistic or uncertain about the future state of the economy, causing investment to decline. The federal government attempted to correct this aggregate demand shock through small business loans, direct aid to state and local governments, expanded unemployment insurance, and stimulus checks. As a result of all these measures, the economy was able to bounce back somewhat over the remainder of 2020. However, even at the start of 2022, millions of people remained out of work as new variants threatened to upend the economy once again.

The Keynesian perspective would have economic policy continue to focus on aggregate and restoring confidence in the economy. President Biden's proposals largely reflect these goals, but with millions of workers still out of the labor market and a virus that is still not contained, it remains to be seen whether those policy prescriptions will be enough "medicine" to bring the economy back to normalcy.

Key Terms

contractionary fiscal policy tax increases or cuts in government spending designed to decrease aggregate demand and reduce inflationary pressures

coordination argument downward wage and price flexibility requires perfect information about the level of lower compensation acceptable to other laborers and market participants

disposable income income after taxes

expansionary fiscal policy tax cuts or increases in government spending designed to stimulate aggregate demand and move the economy out of recession

expenditure multiplier Keynesian concept that asserts that a change in autonomous spending causes a more than proportionate change in real GDP

inflationary gap equilibrium at a level of output above potential GDP

macroeconomic externality occurs when what happens at the macro level is different from and inferior to what happens at the micro level; an example would be where upward sloping supply curves for firms become a flat aggregate supply curve, illustrating that the price level cannot fall to stimulate aggregate demand

menu costs costs firms face in changing prices

Phillips curve the tradeoff between unemployment and inflation

real GDP the amount of goods and services actually sold in a nation

recessionary gap equilibrium at a level of output below potential GDP

sticky wages and prices a situation where wages and prices do not fall in response to a decrease in demand, or do not rise in response to an increase in demand

Key Concepts and Summary

25.1 Aggregate Demand in Keynesian Analysis

Aggregate demand is the sum of four components: consumption, investment, government spending, and net exports. Consumption will change for a number of reasons, including movements in income, taxes, expectations about future income, and changes in wealth levels. Investment will change in response to its expected profitability, which in turn is shaped by expectations about future economic growth, the creation of new technologies, the price of key inputs, and tax incentives for investment. Investment will also change when interest rates rise or fall. Political considerations determine government spending and taxes. Exports and imports change according to relative growth rates and prices between two economies.

25.2 The Building Blocks of Keynesian Analysis

Keynesian economics is based on two main ideas: (1) aggregate demand is more likely than aggregate supply to be the primary cause of a short-run economic event like a recession; (2) wages and prices can be sticky, and so, in an economic downturn, unemployment can result. The latter is an example of a macroeconomic externality. While surpluses cause prices to fall at the micro level, they do not necessarily at the macro level. Instead the adjustment to a decrease in demand occurs only through decreased quantities. One reason why prices may be sticky is menu costs, the costs of changing prices. These include internal costs a business faces in changing prices in terms of labeling, recordkeeping, and accounting, and also the costs of communicating the price change to (possibly unhappy) customers. Keynesians also believe in the existence of the expenditure multiplier—the notion that a change in autonomous expenditure causes a more than proportionate change in GDP.

25.3 The Phillips Curve

A Phillips curve shows the tradeoff between unemployment and inflation in an economy. From a Keynesian viewpoint, the Phillips curve should slope down so that higher unemployment means lower inflation, and vice versa. However, a downward-sloping Phillips curve is a short-term relationship that may shift after a few years.

Keynesian macroeconomics argues that the solution to a recession is expansionary fiscal policy, such as tax cuts to stimulate consumption and investment, or direct increases in government spending that would shift the aggregate demand curve to the right. The other side of Keynesian policy occurs when the economy is operating above potential GDP. In this situation, unemployment is low, but inflationary rises in the price level are a concern. The Keynesian response would be contractionary fiscal policy, using tax increases or government spending cuts to shift AD to the left.

25.4 The Keynesian Perspective on Market Forces

The Keynesian prescription for stabilizing the economy implies government intervention at the macroeconomic level—increasing aggregate demand when private demand falls and decreasing aggregate demand when private demand rises. This does not imply that the government should be passing laws or regulations that set prices and quantities in microeconomic markets.

Self-Check Questions

1. In the Keynesian framework, which of the following events might cause a recession? Which might cause inflation? Sketch AD/AS diagrams to illustrate your answers.
 - a. A large increase in the price of the homes people own.
 - b. Rapid growth in the economy of a major trading partner.
 - c. The development of a major new technology offers profitable opportunities for business.
 - d. The interest rate rises.
 - e. The good imported from a major trading partner become much less expensive.
2. In a Keynesian framework, using an AD/AS diagram, which of the following government policy choices offer a possible solution to recession? Which offer a possible solution to inflation?
 - a. A tax increase on consumer income.
 - b. A surge in military spending.
 - c. A reduction in taxes for businesses that increase investment.
 - d. A major increase in what the U.S. government spends on healthcare.
3. Use the AD/AS model to explain how an inflationary gap occurs, beginning from the initial equilibrium in [Figure 25.6](#).
4. Suppose the U.S. Congress cuts federal government spending in order to balance the Federal budget. Use the AD/AS model to analyze the likely impact on output and employment. *Hint:* revisit [Figure 25.6](#).
5. How would a decrease in energy prices affect the Phillips curve?
6. Does Keynesian economics require government to set controls on prices, wages, or interest rates?
7. List three practical problems with the Keynesian perspective.

Review Questions

8. Name some economic events not related to government policy that could cause aggregate demand to shift.
9. Name some government policies that could cause aggregate demand to shift.
10. From a Keynesian point of view, which is more likely to cause a recession: aggregate demand or aggregate supply, and why?
11. Why do sticky wages and prices increase the impact of an economic downturn on unemployment and recession?
12. Explain what economists mean by “menu costs.”
13. What tradeoff does a Phillips curve show?

14. Would you expect to see long-run data trace out a stable downward-sloping Phillips curve?
15. What is the Keynesian prescription for recession? For inflation?
16. How did the Keynesian perspective address the economic market failure of the Great Depression?

Critical Thinking Questions

17. In its recent report, The Conference Board's *Global Economic Outlook 2015*, updated November 2014 (<http://www.conference-board.org/data/globaloutlook.cfm>), projects China's growth between 2015 and 2019 to be about 5.5%. *International Business Times* (<http://www.ibtimes.com/us-exports-china-have-grown-294-over-past-decade-1338693>) reports that China is the United States' third largest export market, with exports to China growing 294% over the last ten years. Explain what impact China has on the U.S. economy.
18. What may happen if growth in China continues or contracts?
19. Does it make sense that wages would be sticky downwards but not upwards? Why or why not?
20. Suppose the economy is operating at potential GDP when it experiences an increase in export demand. How might the economy increase production of exports to meet this demand, given that the economy is already at full employment?
21. Do you think the Phillips curve is a useful tool for analyzing the economy today? Why or why not?
22. Return to the table from the *Economic Report of the President* in the earlier [Work It Out](#) feature titled "The Phillips Curve for the United States." How would you expect government spending to have changed over the last six years?
23. Explain what types of policies the federal government may have implemented to restore aggregate demand and the potential obstacles policymakers may have encountered.



FIGURE 26.1 Impact of the Great Recession We can see the impact of the Great Recession in many areas of the economy that impact our daily lives. One of the most visible signs was in the housing market where many people were forced to abandon their homes and other buildings, including ones midway through construction. (Credit: modification of "House" by A McLin/Flickr Creative Commons, CC BY 2.0)

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- The Building Blocks of Neoclassical Analysis
- The Policy Implications of the Neoclassical Perspective
- Balancing Keynesian and Neoclassical Models

Introduction to the Neoclassical Perspective



BRING IT HOME

Navigating Uncharted Waters

The Great Recession ended in June 2009 after 18 months, according to the National Bureau of Economic Research (NBER). The NBER examines a variety of measures of economic activity to gauge the economy's overall health. These measures include real income, wholesale and retail sales, employment, and industrial production. In the years since the official end of this historic economic downturn, it has become clear that the Great Recession was two-pronged, hitting the U.S. economy with the collapse of the housing market and the failure of the financial system's credit institutions, further contaminating global economies. While the stock market rapidly lost trillions of dollars of value, consumer spending dried up, and companies began cutting jobs, economic policymakers were struggling with how to best combat and prevent a national, and even global economic collapse. In the end, policymakers used a number of controversial monetary and fiscal policies to support the housing market and domestic industries as well as to stabilize the financial sector. Some of these initiatives included:

- Federal Reserve Bank purchase of both traditional and nontraditional assets off banks' balance sheets. By doing this, the Fed injected money into the banking system and increased the amounts of funds available to lend to the business sector and consumers. This also dropped short-term interest rates to as low as zero percent, which had the effect of devaluing U.S. dollars in the global market and boosting exports.
- The Congress and the President also passed several pieces of legislation that would stabilize the financial market. The Troubled Asset Relief Program (TARP), passed in late 2008, allowed the government to inject cash into troubled banks and other financial institutions and help support General Motors and Chrysler as they faced bankruptcy and threatened job losses throughout their supply chain. The American Recovery and Reinvestment Act in early 2009 provided tax rebates to low- and middle-income households to encourage consumer spending.

Four years after the end of the Great Recession, the economy had yet to return to its pre-recession levels of productivity and growth. Annual productivity increased only 1.9% between 2009 and 2012 compared to its 2.7% annual growth rate between 2000 and 2007, unemployment remained above the natural rate, and real GDP continued to lag behind potential growth. The actions the government took to stabilize the economy were under scrutiny and debate about their effectiveness continues. In this chapter, we will discuss the neoclassical perspective on economics and compare it to the Keynesian perspective, using both the Great Recession and the more recent pandemic-induced recession as examples.

In Chicago, Illinois, the highest recorded temperature was 105° in July 1995, while the lowest recorded temperature was 27° below zero in January 1958. Understanding why these extreme weather patterns occurred would be interesting. However, if you wanted to understand the typical weather pattern in Chicago, instead of focusing on one-time extremes, you would need to look at the entire pattern of data over time.

A similar lesson applies to the study of macroeconomics. It is interesting to study extreme situations, like the 1930s Great Depression, the 2008–2009 Great Recession, or the pandemic-induced recession of 2020. If you want to understand the whole picture, however, you need to look at the long term. Consider the unemployment rate. The unemployment rate has fluctuated from as low as 3.5% in 1969 to as high as 9.7% in 1982 and 8.1% in 2020. Even as the U.S. unemployment rate rose during recessions and declined during expansions, it kept returning to the general neighborhood of 5.0%. When the nonpartisan Congressional Budget Office carried out its long-range economic forecasts in 2010, it assumed that from 2015 to 2020, after the recession has passed, the unemployment rate would be 5.0%. In February 2020, before the COVID-19 pandemic, the unemployment rate reached a historic low of 3.5% and is back to below 5% as of early 2022. From a long-run perspective, the economy seems to keep adjusting back to this rate of unemployment.

As the name “neoclassical” implies, this perspective of how the macroeconomy works is a “new” view of the “old” classical model of the economy. The classical view, the predominant economic philosophy until the Great Depression, was that short-term fluctuations in economic activity would rather quickly, with flexible prices, adjust back to full employment. This view of the economy implied a vertical aggregate supply curve at full employment GDP, and prescribed a “hands off” policy approach. For example, if the economy were to slip into recession (a leftward shift of the aggregate demand curve), it would temporarily exhibit a surplus of goods. Falling prices would eliminate this surplus, and the economy would return to full employment level of GDP. No active fiscal or monetary policy was needed. In fact, the classical view was that expansionary fiscal or monetary policy would only cause inflation, rather than increase GDP. The deep and lasting impact of the Great Depression changed this thinking and Keynesian economics, which prescribed active fiscal policy to alleviate weak aggregate demand, became the more mainstream perspective.

26.1 The Building Blocks of Neoclassical Analysis

LEARNING OBJECTIVES

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

- Explain the importance of potential GDP in the long run
- Analyze the role of flexible prices
- Interpret a neoclassical model of aggregate demand and aggregate supply
- Evaluate different ways for measuring the speed of macroeconomic adjustment

The **neoclassical perspective** on macroeconomics holds that, in the long run, the economy will fluctuate around its potential GDP and its natural rate of unemployment. This chapter begins with two building blocks of neoclassical economics: (1) potential GDP determines the economy's size and (2) wages and prices will adjust in a flexible manner so that the economy will adjust back to its potential GDP level of output. The key policy implication is this: The government should focus more on long-term growth and on controlling inflation than on worrying about recession or cyclical unemployment. This focus on long-run growth rather than the short-run fluctuations in the business cycle means that neoclassical economics is more useful for long-run macroeconomic analysis and Keynesian economics is more useful for analyzing the macroeconomic short run. Let's consider the two neoclassical building blocks in turn, and how we can embody them in the aggregate demand/aggregate supply model.

The Importance of Potential GDP in the Long Run

Over the long run, the level of potential GDP determines the size of real GDP. When economists refer to “potential GDP” they are referring to that level of output that an economy can achieve when all resources (land, labor, capital, and entrepreneurial ability) are fully employed. While the unemployment rate in labor markets will never be zero, full employment in the labor market refers to zero cyclical unemployment. There will still be some level of unemployment due to frictional or structural unemployment, but when the economy is operating with zero cyclical unemployment, economists say that the economy is at the natural rate of unemployment or at full employment.

Economists benchmark actual or real GDP against the potential GDP to determine how well the economy is performing. As explained in [Economic Growth](#), we can explain GDP growth by increases in investment in physical capital and human capital per person as well as advances in technology. **Physical capital per person** refers to the amount and kind of machinery and equipment available to help people get work done. Compare, for example, your productivity in typing a term paper on a typewriter to working on your laptop with word processing software. Clearly, you will be able to be more productive using word processing software. The technology and level of capital of your laptop and software has increased your productivity. More broadly, the development of GPS technology and Universal Product Codes (those barcodes on every product we buy) has made it much easier for firms to track shipments, tabulate inventories, and sell and distribute products. These two technological innovations, and many others, have increased a nation's ability to produce goods and services for a given population. Likewise, increasing human capital involves increasing levels of knowledge, education, and skill sets per person through vocational or higher education. Physical and human capital improvements with technological advances will increase overall productivity and, thus, GDP.

To see how these improvements have increased productivity and output at the national level, we should examine evidence from the United States. The United States experienced significant growth in the twentieth century due to phenomenal changes in infrastructure, equipment, and technological improvements in physical capital and human capital. The population more than tripled in the twentieth century, from 76 million in 1900 to over 300 million in 2016. The human capital of modern workers is far higher today because the education and skills of workers have risen dramatically. In 1900, only about one-eighth of the U.S. population had completed high school and just one person in 40 had completed a four-year college degree. By 2010, about 8.5% of Americans age 25 or older had a high school degree and about 28% had a four-year college degree as well. In 2019, 33% of Americans age 25 or older had a four-year college degree. The average amount of

physical capital per worker has grown dramatically. The technology available to modern workers is extraordinarily better than a century ago: cars, airplanes, electrical machinery, smartphones, computers, chemical and biological advances, materials science, health care—the list of technological advances could run on and on. More workers, higher skill levels, larger amounts of physical capital per worker, and amazingly better technology, and potential GDP for the U.S. economy has clearly increased a great deal since 1900.

This growth has fallen below its potential GDP and, at times, has exceeded its potential. For example from 2008 to 2009, the U.S. economy tumbled into recession and remained below its potential until 2018. After the pandemic-induced recession of March and April 2020, the economy again fell below potential GDP and remains there as of early 2022. At other times, like in the late 1990s or from 2018 to 2020, the economy ran at potential GDP—or even slightly ahead. [Figure 26.2](#) shows the actual data for the increase in real GDP since 1960. The slightly smoother line shows the potential GDP since 1960 as estimated by the nonpartisan Congressional Budget Office. Most economic recessions and upswings are times when the economy is 1–3% below or above potential GDP in a given year. Clearly, short-run fluctuations around potential GDP do exist, but over the long run, the upward trend of potential GDP determines the size of the economy.

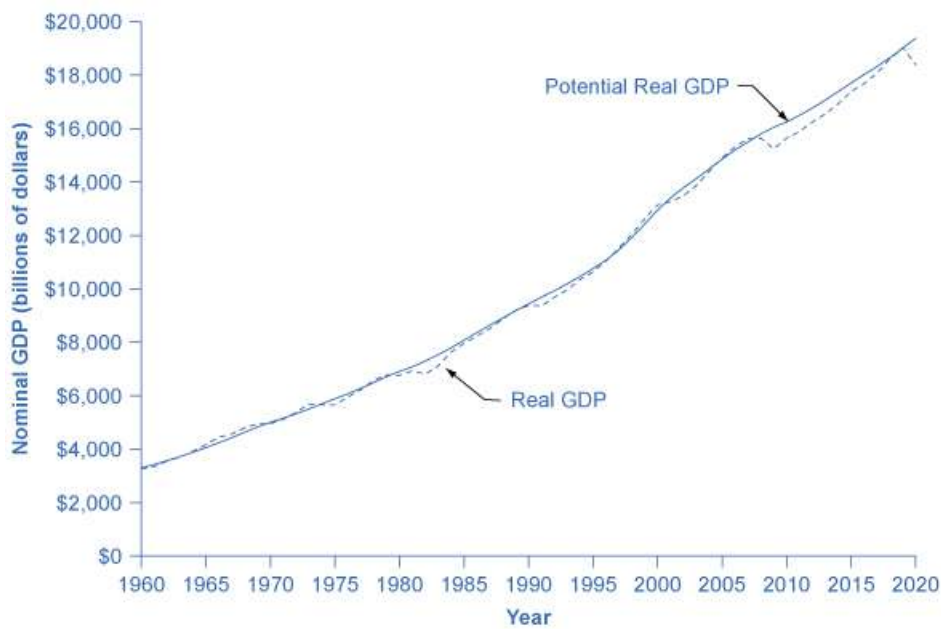


FIGURE 26.2 Potential and Actual GDP (in 2012 Dollars), 1958–2020 Actual GDP falls below potential GDP during and after recessions, like the recessions of 1980 and 1981–82, 1990–91, 2001, and 2008–2009 and continues below potential GDP until 2019, when it goes slightly above potential. In other cases, actual GDP can be above potential GDP for a time, as in the late 1990s.

In the aggregate demand/aggregate supply model, we show potential GDP as a vertical line. Neoclassical economists who focus on potential GDP as the primary determinant of real GDP argue that the long-run aggregate supply curve is located at potential GDP—that is, we draw the long-run aggregate supply curve as a vertical line at the level of potential GDP, as [Figure 26.3](#) shows. A vertical LRAS curve means that the level of aggregate supply (or potential GDP) will determine the economy's real GDP, regardless of the level of aggregate demand. Over time, increases in the quantity and quality of physical capital, increases in human capital, and technological advancements shift potential GDP and the vertical LRAS curve gradually to the right. Economists often describe this gradual increase in an economy's potential GDP as a nation's long-term economic growth.

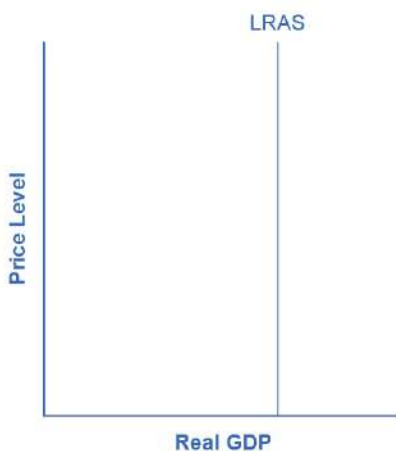


FIGURE 26.3 A Vertical AS Curve In the neoclassical model, we draw the aggregate supply curve as a vertical line at the level of potential GDP. If AS is vertical, then it determines the level of real output, no matter where we draw the aggregate demand curve. Over time, the LRAS curve shifts to the right as productivity increases and potential GDP expands.

The Role of Flexible Prices

How does the macroeconomy adjust back to its level of potential GDP in the long run? What if aggregate demand increases or decreases? Economists base the neoclassical view of how the macroeconomy adjusts on the insight that even if wages and prices are “sticky”, or slow to change, in the short run, they are flexible over time. To understand this better, let's follow the connections from the short-run to the long-run macroeconomic equilibrium.

The aggregate demand and aggregate supply diagram in [Figure 26.4](#) shows two aggregate supply curves. We draw the original upward sloping aggregate supply curve ($SRAS_0$) as a short-run or Keynesian AS curve. The vertical aggregate supply curve ($LRAS_n$) is the long-run or neoclassical AS curve, which is located at potential GDP. The original aggregate demand curve, labeled AD_0 , so that the original equilibrium occurs at point E_0 , at which point the economy is producing at its potential GDP.

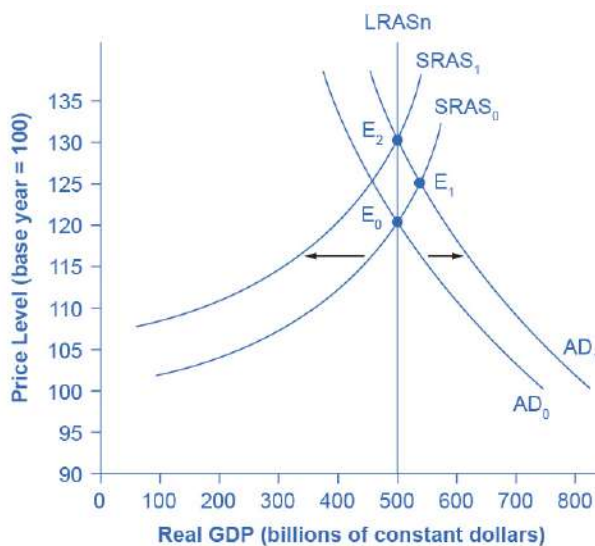


FIGURE 26.4 The Rebound to Potential GDP after AD Increases The original equilibrium (E_0), at an output level of 500 and a price level of 120, happens at the intersection of the aggregate demand curve (AD_0) and the short-run aggregate supply curve ($SRAS_0$). The output at E_0 is equal to potential GDP. Aggregate demand shifts right from AD_0 to AD_1 . The new equilibrium is E_1 , with a higher output level of 550 and an increase in the price level to 125. With

unemployment rates unsustainably low, eager employers bid up wages, which shifts short-run aggregate supply to the left, from $SRAS_0$ to $SRAS_1$. The new equilibrium (E_2) is at the same original level of output, 500, but at a higher price level of 130. Thus, the long-run aggregate supply curve ($LRAS_n$), which is vertical at the level of potential GDP, determines the level of real GDP in this economy in the long run.

Now, imagine that some economic event boosts aggregate demand: perhaps a surge of export sales or a rise in business confidence that leads to more investment, perhaps a policy decision like higher government spending, or perhaps a tax cut that leads to additional aggregate demand. The short-run Keynesian analysis is that the rise in aggregate demand will shift the aggregate demand curve out to the right, from AD_0 to AD_1 , leading to a new equilibrium at point E_1 with higher output, lower unemployment, and pressure for an inflationary rise in the price level.

In the long-run neoclassical analysis, however, the chain of economic events is just beginning. As economic output rises above potential GDP, the level of unemployment falls. The economy is now above full employment and there is a labor shortage. Eager employers are trying to bid workers away from other companies and to encourage their current workers to exert more effort and to work longer hours. This high demand for labor will drive up wages. Most employers review their workers salaries only once or twice a year, and so it will take time before the higher wages filter through the economy. As wages do rise, it will mean a leftward shift in the short-run Keynesian aggregate supply curve back to $SRAS_1$, because the price of a major input to production has increased. The economy moves to a new equilibrium (E_2). The new equilibrium has the same level of real GDP as did the original equilibrium (E_0), but there has been an inflationary increase in the price level.

This description of the short-run shift from E_0 to E_1 and the long-run shift from E_1 to E_2 is a step-by-step way of making a simple point: the economy cannot sustain production above its potential GDP in the long run. An economy may produce above its level of potential GDP in the short run, under pressure from a surge in aggregate demand. Over the long run, however, that surge in aggregate demand ends up as an increase in the price level, not as a rise in output.

The rebound of the economy back to potential GDP also works in response to a shift to the left in aggregate demand. [Figure 26.5](#) again starts with two aggregate supply curves, with $SRAS_0$ showing the original upward sloping short-run Keynesian AS curve and $LRAS_n$ showing the vertical long-run neoclassical aggregate supply curve. A decrease in aggregate demand—for example, because of a decline in consumer confidence that leads to less consumption and more saving—causes the original aggregate demand curve AD_0 to shift back to AD_1 . The shift from the original equilibrium (E_0) to the new equilibrium (E_1) results in a decline in output. The economy is now below full employment and there is a surplus of labor. As output falls below potential GDP, unemployment rises. While a lower price level (i.e., deflation) is rare in the United States, it does happen occasionally during very weak periods of economic activity. For practical purposes, we might consider a lower price level in the AD–AS model as indicative of disinflation, which is a decline in the inflation rate. Thus, the long-run aggregate supply curve $LRAS_n$, which is vertical at the level of potential GDP, ultimately determines this economy's real GDP.

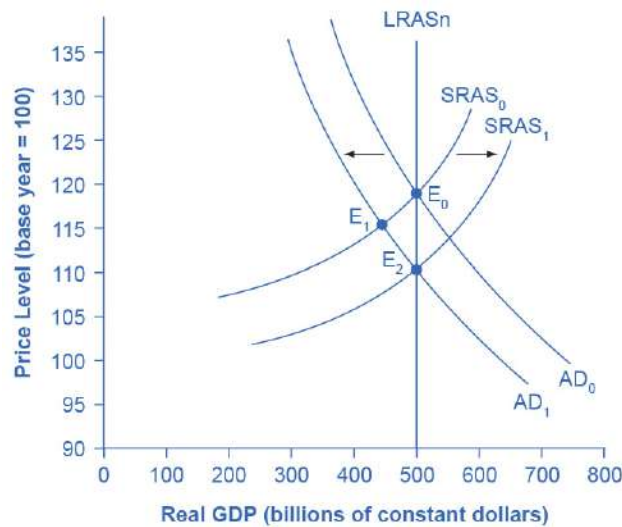


FIGURE 26.5 A Rebound Back to Potential GDP from a Shift to the Left in Aggregate Demand The original equilibrium (E_0), at an output level of 500 and a price level of 120, happens at the intersection of the aggregate demand curve (AD_0) and the short-run aggregate supply curve ($SRAS_0$). The output at E_0 is equal to potential GDP. Aggregate demand shifts left, from AD_0 to AD_1 . The new equilibrium is at E_1 , with a lower output level of 450 and downward pressure on the price level of 115. With high unemployment rates, wages are held down. Lower wages are an economy-wide decrease in the price of a key input, which shifts short-run aggregate supply to the right, from $SRAS_0$ to $SRAS_1$. The new equilibrium (E_2) is at the same original level of output, 500, but at a lower price level of 110.

Again, from the neoclassical perspective, this short-run scenario is only the beginning of the chain of events. The higher level of unemployment means more workers looking for jobs. As a result, employers can hold down on pay increases—or perhaps even replace some of their higher-paid workers with unemployed people willing to accept a lower wage. As wages stagnate or fall, this decline in the price of a key input means that the short-run Keynesian aggregate supply curve shifts to the right from its original ($SRAS_0$ to $SRAS_1$). The overall impact in the long run, as the macroeconomic equilibrium shifts from E_0 to E_1 to E_2 , is that the level of output returns to potential GDP, where it started. There is, however, downward pressure on the price level. Thus, in the neoclassical view, changes in aggregate demand can have a short-run impact on output and on unemployment—but only a short-run impact. In the long run, when wages and prices are flexible, potential GDP and aggregate supply determine real GDP's size.

How Fast Is the Speed of Macroeconomic Adjustment?

How long does it take for wages and prices to adjust, and for the economy to rebound to its potential GDP? This subject is highly contentious. Keynesian economists argue that if the adjustment from recession to potential GDP takes a very long time, then neoclassical theory may be more hypothetical than practical. In response to John Maynard Keynes' immortal words, “In the long run we are all dead,” neoclassical economists respond that even if the adjustment takes as long as, say, ten years the neoclassical perspective remains of central importance in understanding the economy.

One subset of neoclassical economists holds that wage and price adjustment in the macroeconomy might be quite rapid. The theory of **rational expectations** holds that people form the most accurate possible expectations about the future that they can, using all information available to them. In an economy where most people have rational expectations, economic adjustments may happen very quickly.

To understand how rational expectations may affect the speed of price adjustments, think about a situation in the real estate market. Imagine that several events seem likely to push up home values in the neighborhood. Perhaps a local employer announces that it plans to hire many more people or the city announces that it will

build a local park or a library in that neighborhood. The theory of rational expectations points out that even though none of the changes will happen immediately, home prices in the neighborhood will rise immediately, because the expectation that homes will be worth more in the future will lead buyers to be willing to pay more in the present. The amount of the immediate increase in home prices will depend on how likely it seems that the announcements about the future will actually happen and on how distant the local jobs and neighborhood improvements are in the future. The key point is that, because of rational expectations, prices do not wait on events, but adjust immediately.

At a macroeconomic level, the theory of rational expectations points out that if the aggregate supply curve is vertical over time, then people should rationally expect this pattern. When a shift in aggregate demand occurs, people and businesses with rational expectations will know that its impact on output and employment will be temporary, while its impact on the price level will be permanent. If firms and workers perceive the outcome of the process in advance, and if all firms and workers know that everyone else is perceiving the process in the same way, then they have no incentive to go through an extended series of short-run scenarios, like a firm first hiring more people when aggregate demand shifts out and then firing those same people when aggregate supply shifts back. Instead, everyone will recognize where this process is heading—toward a change in the price level—and then will act on that expectation. In this scenario, the expected long-run change in the price level may happen very quickly, without a drawn-out zigzag of output and employment first moving one way and then the other.

The theory that people and firms have rational expectations can be a useful simplification, but as a statement about how people and businesses actually behave, the assumption seems too strong. After all, many people and firms are not especially well informed, either about what is happening in the economy or about how the economy works. An alternate assumption is that people and firms act with **adaptive expectations**: they look at past experience and gradually adapt their beliefs and behavior as circumstances change, but are not perfect synthesizers of information and accurate predictors of the future in the sense of rational expectations theory. If most people and businesses have some form of adaptive expectations, then the adjustment from the short run and long run will be traced out in incremental steps that occur over time.

The empirical evidence on the speed of macroeconomic adjustment of prices and wages is not clear-cut. The speed of macroeconomic adjustment probably varies among different countries and time periods. A reasonable guess is that the initial short-run effect of a shift in aggregate demand might last two to five years, before the adjustments in wages and prices cause the economy to adjust back to potential GDP. Thus, one might think of the short run for applying Keynesian analysis as time periods less than two to five years, and the long run for applying neoclassical analysis as longer than five years. For practical purposes, this guideline is frustratingly imprecise, but when analyzing a complex social mechanism like an economy as it evolves over time, some imprecision seems unavoidable.

26.2 The Policy Implications of the Neoclassical Perspective

LEARNING OBJECTIVES

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

- Discuss why and how economists measure inflation expectations
- Analyze the impacts of fiscal and monetary policy on aggregate supply and aggregate demand
- Explain the neoclassical Phillips curve, noting its tradeoff between inflation and unemployment
- Identify clear distinctions between neoclassical economics and Keynesian economics

To understand the policy recommendations of the neoclassical economists, it helps to start with the Keynesian perspective. Suppose a decrease in aggregate demand causes the economy to go into recession with high unemployment. The Keynesian response would be to use government policy to stimulate aggregate demand and eliminate the recessionary gap. The neoclassical economists believe that the Keynesian response, while perhaps well intentioned, will not have a good outcome for reasons we will discuss shortly. Since the neoclassical economists believe that the economy will correct itself over time, the only advantage of a

Keynesian stabilization policy would be to accelerate the process and minimize the time that the unemployed are out of work. Is that the likely outcome?

Keynesian macroeconomic policy requires some optimism about the government's ability to recognize a situation of too little or too much aggregate demand, and to adjust aggregate demand accordingly with the right level of changes in taxes or spending, all enacted in a timely fashion. After all, neoclassical economists argue, it takes government statisticians months to produce even preliminary estimates of GDP so that politicians know whether a recession is occurring—and those preliminary estimates may be revised substantially later. Moreover, there is the question of timely action. The political process can take more months to enact a tax cut or a spending increase. Political or economic considerations may determine the amount of tax or spending changes. Then the economy will take still more months to put into effect changes in aggregate demand through spending and production. When economists and policy makers consider all of these time lags and political realities, active fiscal policy may fail to address the current problem, and could even make the future economy worse. The average U.S. post-World War II recession has lasted only about a year. By the time government policy activates, the recession will likely be over. As a consequence, the only result of government fine-tuning will be to stimulate the economy when it is already recovering (or to contract the economy when it is already falling). In other words, an active macroeconomic policy is likely to exacerbate the cycles rather than dampen them. Some neoclassical economists believe a large part of the business cycles we observe are due to flawed government policy. To learn about this issue further, read the following Clear It Up feature.



CLEAR IT UP

Why and how do economists measure inflation expectations?

People take expectations about inflation into consideration every time they make a major purchase, such as a house or a car. As inflation fluctuates, so too does the nominal interest rate on loans to buy these goods. The nominal interest rate is comprised of the real rate, plus an **expected inflation** factor. Expected inflation also tells economists about how the public views the economy's direction. Suppose the public expects inflation to increase. This could be the result of positive demand shock due to an expanding economy and increasing aggregate demand. It could also be the result of a negative supply shock, perhaps from rising energy prices, and decreasing aggregate supply. In either case, the public may expect the central bank to engage in contractionary monetary policy to reduce inflation, and this policy results in higher interest rates. If, however economists expect inflation to decrease, the public may anticipate a recession. In turn, the public may expect expansionary monetary policy, and lower interest rates, in the short run. By monitoring expected inflation, economists garner information about the effectiveness of macroeconomic policies. Additionally, monitoring expected inflation allows for projecting the direction of real interest rates that isolate for the effect of inflation. This information is necessary for making decisions about financing investments.

Expectations about inflation may seem like a highly theoretical concept, but, in fact the Federal Reserve Bank measures, inflation expectations based upon early research conducted by Joseph Livingston, a financial journalist for the *Philadelphia Inquirer*. In 1946, he started a twice-a-year survey of economists about their expectations of inflation. After Livingston's death in 1969, the Federal Reserve Bank and other economic research agencies such as the Survey Research Center at the University of Michigan, the American Statistical Association, and the National Bureau of Economic Research continued the survey.

Current Federal Reserve research compares these expectations to actual inflation that has occurred, and the results, so far, are mixed. Economists' forecasts, however, have become notably more accurate in the last few decades. Economists are actively researching how inflation expectations and other economic variables form and change.

LINK IT UP

Visit this [website \(https://www.clevelandfed.org/newsroom-and-events/publications/economic-commentary/economic-commentary-archives/2009-economic-commentaries/ec-20090809-a-new-approach-to-gauging-inflation-expectations.aspx\)](https://www.clevelandfed.org/newsroom-and-events/publications/economic-commentary/economic-commentary-archives/2009-economic-commentaries/ec-20090809-a-new-approach-to-gauging-inflation-expectations.aspx) to read “The Federal Reserve Bank of Cleveland’s Economic Commentary: A New Approach to Gauging Inflation Expectations” by Joseph G. Haubrich for more information about how economists forecast expected inflation.

The Neoclassical Phillips Curve Tradeoff

The [Keynesian Perspective](#) introduced the Phillips curve and explained how it is derived from the aggregate supply curve. The short run upward sloping aggregate supply curve implies a downward sloping Phillips curve; thus, there is a tradeoff between inflation and unemployment in the short run. By contrast, a neoclassical long-run aggregate supply curve will imply a vertical shape for the Phillips curve, indicating no long run tradeoff between inflation and unemployment. [Figure 26.6](#) (a) shows the vertical AS curve, with three different levels of aggregate demand, resulting in three different equilibria, at three different price levels. At every point along that vertical AS curve, potential GDP and the rate of unemployment remains the same. Assume that for this economy, the natural rate of unemployment is 5%. As a result, the long-run Phillips curve relationship, in [Figure 26.6](#) (b), is a vertical line, rising up from 5% unemployment, at any level of inflation. Read the following Work It Out feature for additional information on how to interpret inflation and unemployment rates.

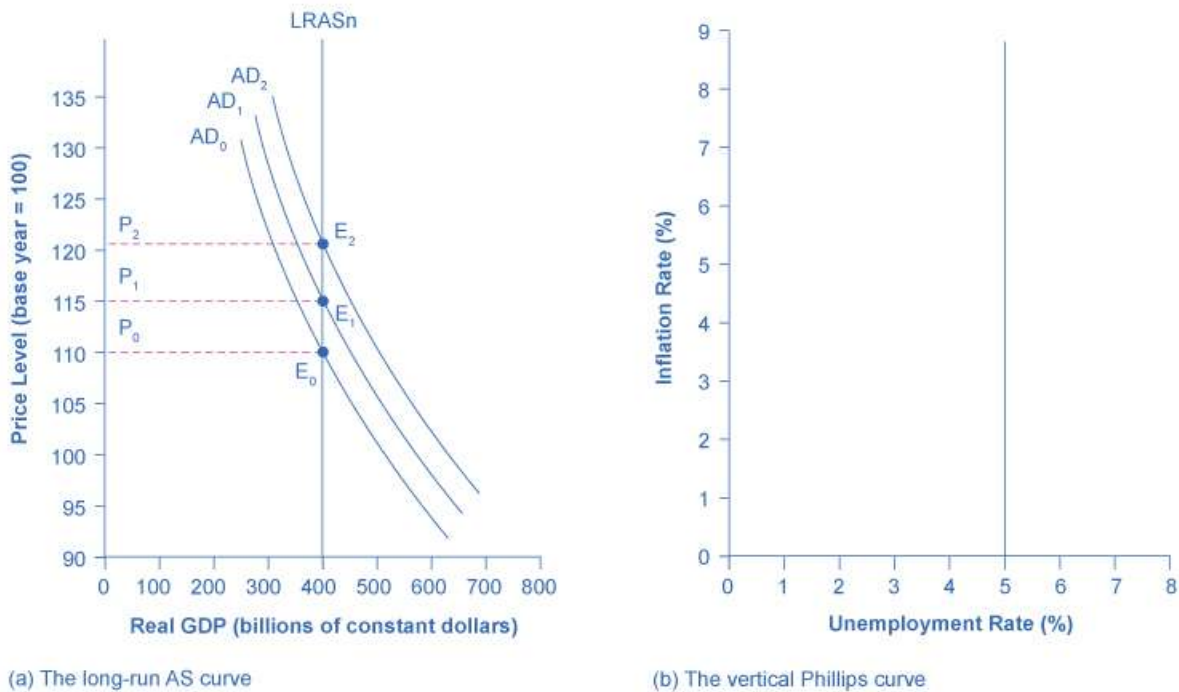


FIGURE 26.6 From a Long-Run AS Curve to a Long-Run Phillips Curve (a) With a vertical LRAS curve, shifts in aggregate demand do not alter the level of output but do lead to changes in the price level. Because output is unchanged between the equilibria E_0 , E_1 , and E_2 , all unemployment in this economy will be due to the natural rate of unemployment. (b) If the natural rate of unemployment is 5%, then the Phillips curve will be vertical. That is, regardless of changes in the price level, the unemployment rate remains at 5%.

WORK IT OUT

Tracking Inflation and Unemployment Rates

Suppose that you have collected data for years on inflation and unemployment rates and recorded them in a table, such as [Table 26.1](#). How do you interpret that information?

Year	Inflation Rate	Unemployment Rate
1970	2%	4%
1975	3%	3%
1980	2%	4%
1985	1%	6%
1990	1%	4%
1995	4%	2%
2000	5%	4%

TABLE 26.1

Step 1. Plot the data points in a graph with inflation rate on the vertical axis and unemployment rate on the horizontal axis. Your graph will appear similar to [Figure 26.7](#).

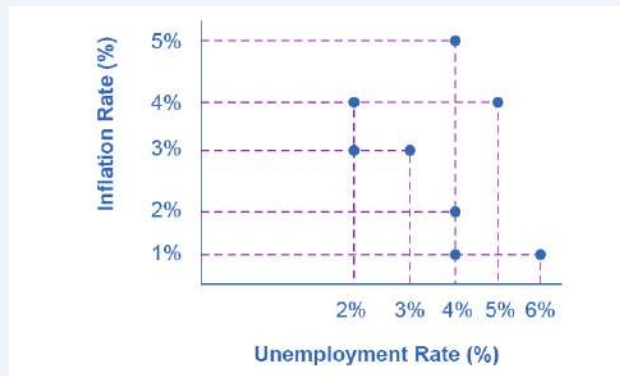


FIGURE 26.7 Inflation Rates

Step 2. What patterns do you see in the data? You should notice that there are years when unemployment falls but inflation rises, and other years where unemployment rises and inflation falls.

Step 3. Can you determine the natural rate of unemployment from the data or from the graph? As you analyze the graph, it appears that the natural rate of unemployment lies at 4%. This is the rate that the economy appears to adjust back to after an apparent change in the economy. For example, in 1975 the economy appeared to have an increase in aggregate demand. The unemployment rate fell to 3% but inflation increased from 2% to 3%. By 1980, the economy had adjusted back to 4% unemployment and the inflation rate had returned to 2%. In 1985, the economy looks to have suffered a recession as unemployment rose to 6% and inflation fell to 1%. This would be consistent with a decrease in aggregate demand. By 1990, the economy recovered back to 4% unemployment, but at a lower inflation rate of 1%. In 1995 the economy again rebounded and unemployment fell to 2%, but inflation increased to 4%, which is consistent with a large increase in aggregate demand. The

economy adjusted back to 4% unemployment but at a higher rate of inflation of 5%. Then in 2000, both unemployment and inflation increased to 5% and 4%, respectively.

Step 4. Do you see the Phillips curve(s) in the data? If we trace the downward sloping trend of data points, we could see a short-run Phillips curve that exhibits the inverse tradeoff between higher unemployment and lower inflation rates. If we trace the vertical line of data points, we could see a long-run Phillips curve at the 4% natural rate of unemployment.

The unemployment rate on the long-run Phillips curve will be the natural rate of unemployment. A small inflationary increase in the price level from AD_0 to AD_1 will have the same natural rate of unemployment as a larger inflationary increase in the price level from AD_0 to AD_2 . The macroeconomic equilibrium along the vertical aggregate supply curve can occur at a variety of different price levels, and the natural rate of unemployment can be consistent with all different rates of inflation. The great economist Milton Friedman (1912–2006) summed up the neoclassical view of the long-term Phillips curve tradeoff in a 1967 speech: “[T]here is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off.”

In the Keynesian perspective, the primary focus is on getting the level of aggregate demand right in relationship to an upward-sloping aggregate supply curve. That is, the government should adjust AD so that the economy produces at its potential GDP, not so low that cyclical unemployment results and not so high that inflation results. In the neoclassical perspective, aggregate supply will determine output at potential GDP, the natural rate of unemployment determines unemployment, and shifts in aggregate demand are the primary determinant of changes in the price level.

LINK IT UP

Visit this [website \(http://openstax.org/l/modeledbehavior\)](http://openstax.org/l/modeledbehavior) to read about the effects of economic intervention.

Fighting Unemployment or Inflation?

As we explained in [Unemployment](#), economists divide unemployment into two categories: cyclical unemployment and the natural rate of unemployment, which is the sum of frictional and structural unemployment. Cyclical unemployment results from fluctuations in the business cycle and is created when the economy is producing below potential GDP—giving potential employers less incentive to hire. When the economy is producing at potential GDP, cyclical unemployment will be zero. Because of labor market dynamics, in which people are always entering or exiting the labor force, the unemployment rate never falls to 0%, not even when the economy is producing at or even slightly above potential GDP. Probably the best we can hope for is for the number of job vacancies to equal the number of job seekers. We know that it takes time for job seekers and employers to find each other, and this time is the cause of frictional unemployment. Most economists do not consider frictional unemployment to be a “bad” thing. After all, there will always be workers who are unemployed while looking for a job that is a better match for their skills. There will always be employers that have an open position, while looking for a worker that is a better match for the job. Ideally, these matches happen quickly, but even when the economy is very strong there will be some natural unemployment and this is what the natural rate of unemployment measures.

The neoclassical view of unemployment tends to focus attention away from the cyclical unemployment problem—that is, unemployment caused by recession—while putting more attention on the unemployment rate issue that prevails even when the economy is operating at potential GDP. To put it another way, the neoclassical view of unemployment tends to focus on how the government can adjust public policy to reduce the natural rate of unemployment. Such policy changes might involve redesigning unemployment and welfare programs so that they support those in need, but also offer greater encouragement for job-hunting. It might

involve redesigning business rules with an eye to whether they are unintentionally discouraging businesses from taking on new employees. It might involve building institutions to improve the flow of information about jobs and the mobility of workers, to help bring workers and employers together more quickly. For those workers who find that their skills are permanently no longer in demand (for example, the structurally unemployed), economists can design policy to provide opportunities for retraining so that these workers can reenter the labor force and seek employment.

Neoclassical economists will not tend to see aggregate demand as a useful tool for reducing unemployment; after all, with a vertical aggregate supply curve determining economic output, then aggregate demand has no long-run effect on unemployment. Instead, neoclassical economists believe that aggregate demand should be allowed to expand only to match the gradual shifts of aggregate supply to the right—keeping the price level much the same and inflationary pressures low.

If aggregate demand rises rapidly in the neoclassical model, in the long run it leads only to inflationary pressures. [Figure 26.8](#) shows a vertical LRAS curve and three different levels of aggregate demand, rising from AD_0 to AD_1 to AD_2 . As the macroeconomic equilibrium rises from E_0 to E_1 to E_2 , the price level rises, but real GDP does not budge; nor does the rate of unemployment, which adjusts to its natural rate. Conversely, reducing inflation has no long-term costs, either. Think about [Figure 26.8](#) in reverse, as the aggregate demand curve shifts from AD_2 to AD_1 to AD_0 , and the equilibrium moves from E_2 to E_1 to E_0 . During this process, the price level falls, but, in the long run, neither real GDP nor the natural unemployment rate changes.

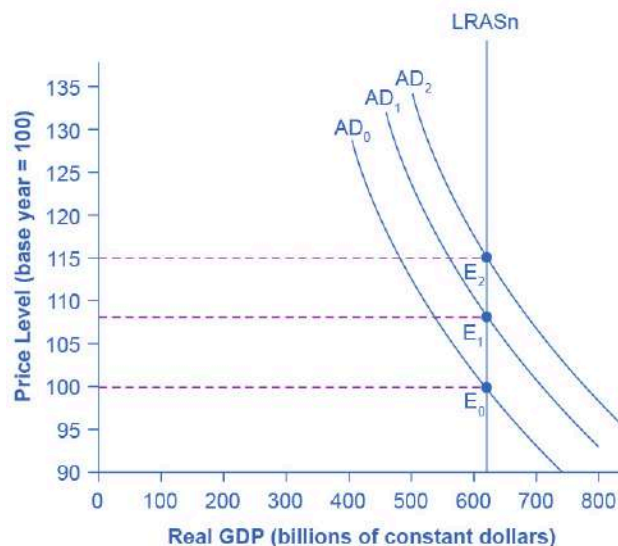


FIGURE 26.8 How Aggregate Demand Determines the Price Level in the Long Run As aggregate demand shifts to the right, from AD_0 to AD_1 to AD_2 , real GDP in this economy and the level of unemployment do not change. However, there is inflationary pressure for a higher price level as the equilibrium changes from E_0 to E_1 to E_2 .

LINK IT UP

Visit this [website \(http://openstax.org/l/inflatemploy\)](http://openstax.org/l/inflatemploy) to read about how inflation and unemployment are related.

Fighting Recession or Encouraging Long-Term Growth?

Neoclassical economists believe that the economy will rebound out of a recession or eventually contract during an expansion because prices and wage rates are flexible and will adjust either upward or downward to restore the economy to its potential GDP. Thus, the key policy question for neoclassicals is how to promote growth of potential GDP. We know that economic growth ultimately depends on the growth rate of long-term productivity.

Productivity measures how effective inputs are at producing outputs. We know that U.S. productivity has grown on average about 2% per year. That means that the same amount of inputs produce 2% more output than the year before. We also know that productivity growth varies a great deal in the short term due to cyclical factors. It also varies somewhat in the long term. From 1953–1972, U.S. labor productivity (as measured by output per hour in the business sector) grew at 3.2% per year. From 1973–1992, productivity growth declined significantly to 1.8% per year. Then, from 1993–2010, productivity growth increased to around 2% per year. In recent years, it has grown less than 2% per year, although it did pick up in 2019 and 2020 to over 2% again. The neoclassical economists believe the underpinnings of long-run productivity growth to be an economy's investments in human capital, physical capital, and technology, operating together in a market-oriented environment that rewards innovation. Government policy should focus on promoting these factors.

Summary of Neoclassical Macroeconomic Policy Recommendations

Let's summarize what neoclassical economists recommend for macroeconomic policy. Neoclassical economists do not believe in “fine-tuning” the economy. They believe that a stable economic environment with a low rate of inflation fosters economic growth. Similarly, tax rates should be low and unchanging. In this environment, private economic agents can make the best possible investment decisions, which will lead to optimal investment in physical and human capital as well as research and development to promote improvements in technology.

Summary of Neoclassical Economics versus Keynesian Economics

[Table 26.2](#) summarizes the key differences between the two schools of thought.

Summary	Neoclassical Economics	Keynesian Economics
Focus: long-term or short term	Long-term	Short-term
Prices and wages: sticky or flexible?	Flexible	Sticky
Economic output: Primarily determined by aggregate demand or aggregate supply?	Aggregate supply	Aggregate demand
Aggregate supply: vertical or upward-sloping?	Vertical	Upward-sloping
Phillips curve vertical or downward-sloping	Vertical	Downward sloping
Is aggregate demand a useful tool for controlling inflation?	Yes	Yes
What should be the primary area of policy emphasis for reducing unemployment?	Reform labor market institutions to reduce natural rate of unemployment	Increase aggregate demand to eliminate cyclical unemployment
Is aggregate demand a useful tool for ending recession?	At best, only in the short-run temporary sense, but may just increase inflation instead	Yes

TABLE 26.2 Neoclassical versus Keynesian Economics

26.3 Balancing Keynesian and Neoclassical Models

LEARNING OBJECTIVES

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

- Evaluate how neoclassical economists and Keynesian economists react to recessions
- Analyze the interrelationship between the neoclassical and Keynesian economic models

We can compare finding the balance between Keynesian and Neoclassical models to the challenge of riding two horses simultaneously. When a circus performer stands on two horses, with a foot on each one, much of the excitement for the viewer lies in contemplating the gap between the two. As modern macroeconomists ride into the future on two horses—with one foot on the short-term Keynesian perspective and one foot on the long-term neoclassical perspective—the balancing act may look uncomfortable, but there does not seem to be any way to avoid it. Each approach, Keynesian and neoclassical, has its strengths and weaknesses.

The short-term Keynesian model, built on the importance of aggregate demand as a cause of business cycles and a degree of wage and price rigidity, does a sound job of explaining many recessions and why cyclical unemployment rises and falls. By focusing on the short-run aggregate demand adjustments, Keynesian economics risks overlooking the long-term causes of economic growth or the natural rate of unemployment that exist even when the economy is producing at potential GDP.

The neoclassical model, with its emphasis on aggregate supply, focuses on the underlying determinants of output and employment in markets, and thus tends to put more emphasis on economic growth and how labor markets work. However, the neoclassical view is not especially helpful in explaining why unemployment moves up and down over short time horizons of a few years. Nor is the neoclassical model especially helpful when the economy is mired in an especially deep and long-lasting recession, like the 1930s Great Depression. Keynesian economics tends to view inflation as a price that might sometimes be paid for lower unemployment; neoclassical economics tends to view inflation as a cost that offers no offsetting gains in terms of lower unemployment.

Macroeconomics cannot, however, be summed up as an argument between one group of economists who are pure Keynesians and another group who are pure neoclassicists. Instead, many mainstream economists believe both the Keynesian and neoclassical perspectives. Robert Solow, the Nobel laureate in economics in 1987, described the dual approach in this way:

At short time scales, I think, something sort of ‘Keynesian’ is a good approximation, and surely better than anything straight ‘neoclassical.’ At very long time scales, the interesting questions are best studied in a neoclassical framework, and attention to the Keynesian side of things would be a minor distraction. At the five-to-ten-year time scale, we have to piece things together as best we can, and look for a hybrid model that will do the job.

Many modern macroeconomists spend considerable time and energy trying to construct models that blend the most attractive aspects of the Keynesian and neoclassical approaches. It is possible to construct a somewhat complex mathematical model where aggregate demand and sticky wages and prices matter in the short run, but wages, prices, and aggregate supply adjust in the long run. However, creating an overall model that encompasses both short-term Keynesian and long-term neoclassical models is not easy.



BRING IT HOME

Navigating Uncharted Waters—The Great Recession and Pandemic-Induced Recession of 2020

Were the policies that the government implemented to stabilize the economy and financial markets during the Great Recession of 2007–2009, and the pandemic-induced recession of 2020 effective? Many economists from both the Keynesian and neoclassical schools have found that they were, although to varying degrees. Regarding the Great Recession, Alan Blinder of Princeton University and Mark Zandi for Moody’s Analytics found that, without fiscal

policy, GDP decline would have been significantly more than its 3.3% in 2008 followed by its 0.1% decline in 2009. They also estimated that there would have been 8.5 million more job losses had the government not intervened in the market with the TARP to support the financial industry and key automakers General Motors and Chrysler. Federal Reserve Bank economists Carlos Carvalho, Stefano Eusip, and Christian Grisse found in their study, *Policy Initiatives in the Global Recession: What Did Forecasters Expect?* that once the government implemented policies, forecasters adapted their expectations to these policies. They were more likely to anticipate increases in investment due to lower interest rates brought on by monetary policy and increased economic growth resulting from fiscal policy.

The neoclassical perspective can also shed light on the country's experience with policy during the pandemic-induced recession of 2020. It was mentioned earlier that one criticism made by proponents of the neoclassical perspective is that government policy is often too slow to react to a recession. However after the pandemic hit, the federal government quickly responded with aid to state and local governments, increased unemployment insurance, aid to businesses forced to shut down, and stimulus checks to boost spending. There is no doubt that the economic fallout from the pandemic would have been much worse without these policies. Some economists even argue that the government helped too much and that the high inflation the U.S. economy experienced starting mid-2021 is due to the real output growing faster than potential, but it is too early (as of early 2022) to tell if that argument is correct.

By focusing on potential GDP instead of short-run demand, the neoclassical perspective also makes an important point about how the size of the economy determines its ability to grow. Since the pandemic hit, millions of workers have stayed out of the labor market due to early retirement, health and safety concerns, the availability of childcare, and school closures. As mentioned in [Unemployment](#), these changes have caused labor force participation to remain lower than its historical average. The pandemic has also made it harder for future workers to acquire skills they need to be productive in the labor market. The longer these dynamics are at play, the more harm it will do to potential GDP.

Key Terms

adaptive expectations the theory that people look at past experience and gradually adapt their beliefs and behavior as circumstances change

expected inflation a future rate of inflation that consumers and firms build into current decision making

neoclassical perspective the philosophy that, in the long run, the business cycle will fluctuate around the potential, or full-employment, level of output

physical capital per person the amount and kind of machinery and equipment available to help a person produce a good or service

rational expectations the theory that people form the most accurate possible expectations about the future that they can, using all information available to them

Key Concepts and Summary

26.1 The Building Blocks of Neoclassical Analysis

The neoclassical perspective argues that, in the long run, the economy will adjust back to its potential GDP level of output through flexible price levels. Thus, the neoclassical perspective views the long-run AS curve as vertical. A rational expectations perspective argues that people have excellent information about economic events and how the economy works and that, as a result, price and other economic adjustments will happen very quickly. In adaptive expectations theory, people have limited information about economic information and how the economy works, and so price and other economic adjustments can be slow.

26.2 The Policy Implications of the Neoclassical Perspective

Neoclassical economists tend to put relatively more emphasis on long-term growth than on fighting recession, because they believe that recessions will fade in a few years and long-term growth will ultimately determine the standard of living. They tend to focus more on reducing the natural rate of unemployment caused by economic institutions and government policies than the cyclical unemployment caused by recession.

Neoclassical economists also see no social benefit to inflation. With an upward-sloping Keynesian AS curve, inflation can arise because an economy is approaching full employment. With a vertical long-run neoclassical AS curve, inflation does not accompany any rise in output. If aggregate supply is vertical, then aggregate demand does not affect the quantity of output. Instead, aggregate demand can only cause inflationary changes in the price level. A vertical aggregate supply curve, where the quantity of output is consistent with many different price levels, also implies a vertical Phillips curve.

26.3 Balancing Keynesian and Neoclassical Models

The Keynesian perspective considers changes to aggregate demand to be the cause of business cycle fluctuations. Keynesians are likely to advocate that policy makers actively attempt to reverse recessionary and inflationary periods because they are not convinced that the self-correcting economy can easily return to full employment.

The neoclassical perspective places more emphasis on aggregate supply. Neoclassical economists believe that long term productivity growth determines the potential GDP level and that the economy typically will return to full employment after a change in aggregate demand. Skeptical of the effectiveness and timeliness of Keynesian policy, neoclassical economists are more likely to advocate a hands-off, or fairly limited, role for active stabilization policy.

While Keynesians would tend to advocate an acceptable tradeoff between inflation and unemployment when counteracting a recession, neoclassical economists argue that no such tradeoff exists. Any short-term gains in lower unemployment will eventually vanish and the result of active policy will only be inflation.

Self-Check Questions

1. Do rational expectations tend to look back at past experience while adaptive expectations look ahead to the future? Explain your answer.
2. Legislation proposes that the government should use macroeconomic policy to achieve an unemployment rate of zero percent, by increasing aggregate demand for as much and as long as necessary to accomplish this goal. From a neoclassical perspective, how will this policy affect output and the price level in the short run and in the long run? Sketch an aggregate demand/aggregate supply diagram to illustrate your answer. *Hint: revisit [Figure 26.4](#).*
3. Would it make sense to argue that rational expectations economics is an extreme version of neoclassical economics? Explain.
4. Summarize the Keynesian and Neoclassical models.

Review Questions

5. Does neoclassical economics focus on the long term or the short term? Explain your answer.
6. Does neoclassical economics view prices and wages as sticky or flexible? Why?
7. What shape is the long-run aggregate supply curve? Why does it have this shape?
8. What is the difference between rational expectations and adaptive expectations?
9. A neoclassical economist and a Keynesian economist are studying the economy of Vineland. It appears that Vineland is beginning to experience a mild recession with a decrease in aggregate demand. Which of these two economists would likely advocate that the government of Vineland take active measures to reverse this decline in aggregate demand? Why?
10. Do neoclassical economists tend to focus more on long term economic growth or on recessions? Explain briefly.
11. Do neoclassical economists tend to focus more on cyclical unemployment or on inflation? Explain briefly.
12. Do neoclassical economists see a value in tolerating a little more inflation if it brings additional economic output? Explain your answer.
13. If aggregate supply is vertical, what role does aggregate demand play in determining output? In determining the price level?
14. What is the shape of the neoclassical long-run Phillips curve? What assumptions do economists make that lead to this shape?
15. When the economy is experiencing a recession, why would a neoclassical economist be unlikely to argue for aggressive policy to stimulate aggregate demand and return the economy to full employment? Explain your answer.
16. If the economy is suffering through a rampant inflationary period, would a Keynesian economist advocate for stabilization policy that involves higher taxes and higher interest rates? Explain your answer.

Critical Thinking Questions

17. If most people have rational expectations, how long will recessions last?
18. Explain why the neoclassical economists believe that the government does not need to do much about unemployment. Do you agree or disagree? Explain.

19. Economists from all theoretical persuasions criticized the American Recovery and Reinvestment Act. The “Stimulus Package” was arguably a Keynesian measure so why would a Keynesian economist be critical of it? Why would neoclassical economists be critical?
20. Is it a logical contradiction to be a neoclassical Keynesian? Explain.

Problems

21. Use [Table 26.3](#) to answer the following questions.

Price Level	Aggregate Supply	Aggregate Demand
90	3,000	3,500
95	3,000	3,000
100	3,000	2,500
105	3,000	2,200
110	3,000	2,100

TABLE 26.3

- Sketch an aggregate supply and aggregate demand diagram.
- What is the equilibrium output and price level?
- If aggregate demand shifts right, what is equilibrium output?
- If aggregate demand shifts left, what is equilibrium output?
- In this scenario, would you suggest using aggregate demand to alter the level of output or to control any inflationary increases in the price level?