

# Macroeconomics Made Clear

Course Guidebook

Akila Weerapana



## LEADERSHIP

---

President & CEO	PAUL SUIJK
Chief Financial Officer	BRUCE G. WILLIS
Chief Marketing Officer	CALE PRITCHETT
SVP, Marketing	JOSEPH PECKL
SVP, Content Development	JASON SMIGEL
VP, Content Production	KEVIN BARNHILL
VP, Marketing	EMILY COOPER
VP, Customer Engagement	KONSTANTINE GELFOND
VP, Technology Services	MARK LEONARD
VP, Content Strategy	KEVIN MANZEL
VP, People	AUDREY WILLIAMS
General Counsel	DEBRA STORMS
Sr. Director, Content Operations	GAIL GLEESON
Director, Talent Acquisition	WILLIAM SCHMIDT
Director, Creative	OCTAVIA VANNALL

## PRODUCTION

---

Studio Operations Manager	JIM M. ALLEN
Video Production Director	ROBERTO DE MORAES
Technical Engineering Manager	SAL RODRIGUEZ
Quality Assurance Supervisor	JAMIE MCCOMBER
Sr. Postproduction Manager	PETER DWYER
Sr. Manager, Production	RIMA KHALEK
Executive Producer	JAY TATE
Producer	TRISH GOLDEN
Studio Producer	INES I. PEREZ-THOMPSON
Sr. Image Rights Analyst	LISA PERSINGER ROBERTSON
Postproduction Manager	OWEN YOUNG
Sr. Editor	MILES MCNAMEE
Sr. Audio Engineer	ED SALTZMAN
Camera Operators	LAKE MANNIKKO
	VALERIE WELCH
Production Assistant	PAUL SHEEHAN

## EDITORIAL & DESIGN SERVICES

---

Director	FARHAD HOSSAIN
Sr. Managing Editor	BLAKELY SWAIN
Editorial Associates	MOLLY LEVY
	MARGI WILHELM
Editorial Assistant	WILLIAM DOMANSKI
Graphics Manager	JAMES NIDEL
Sr. Graphic Artist	BRIAN SCHUMACHER
Graphics Coordinator	KATE STEINBAUER
Graphic Designer	RHOCHELLE MUNSAYAC
Editorial Services	PROOFED



## Akila Weerapana

**Akila Weerapana** is an Associate Professor of Economics at Wellesley College who earned a PhD in Economics from Stanford University. He has won Wellesley College's Pinanski Teaching Prize and appeared in The Princeton Review's *Best 300 Professors*. He is a coauthor of the textbook *Principles of Economics* and has published work in the *Journal of Economic Perspectives*, *The Journal of Economic Education*, and *Economic Development and Cultural Change*, among other publications.



# Table of Contents

About Akila Weerapana	i
<b>1</b> Viewing the World through a New Lens	1
<b>2</b> GDP: The Economy in a Single Number?	11
<b>3</b> Prices: How to Quantify the Dangers of Inflation	20
<b>4</b> Unemployment: Who Wants to Work?	30
<b>5</b> How Is Money Created—and Why?	41
<b>6</b> Is There Such a Thing as Too Much Money?	51
<b>7</b> Two Interest Rates to Rule Them All?	60
<b>8</b> Why Economic Growth Is So Powerful—and Hard	69
<b>9</b> Technology: The Engine of Sustained Growth	79
<b>10</b> What Are Fundamental Barriers to Growth?	88
<b>11</b> How Recessions Occur and Why They Matter	99
<b>12</b> An Intuitive Model of Economic Recessions	109
<b>13</b> Lessons from Past Recessions	119
<b>14</b> What Is Good Monetary or Fiscal Policy?	129
<b>15</b> Extraordinary Actions by Central Banks	138
<b>16</b> How Much Debt Is Too Much for a Country?	148
<b>17</b> How Exchange Rates Move the Global Economy	157
<b>18</b> Why Countries Try to Control Exchange Rates	166
<b>19</b> The Euro: Many Nations under One Currency?	176
<b>20</b> International Trade and Finance Must Balance	186
<b>21</b> How Stocks and the Economy Shape Each Other	195
<b>22</b> Labor Market Policies for the Long Term	204
<b>23</b> Rising Inequality Is a Threat to All of Us	216
<b>24</b> Three Macroeconomic Challenges of Our Time	225

# 1

## VIEWING THE WORLD THROUGH A NEW LENS

**M**acroeconomics seeks answers to significant questions that impact the lives of billions of people. In this course, you will use some simple models, such as the Solow model and the aggregate demand–inflation adjustment model, to shed light on various relationships in the economy. Using macroeconomic models to reach policy conclusions can often lead to disagreements, but it is through such disagreements that the science of economics advances. You will explore important questions related to economic growth, economic fluctuations, and the interplay between macroeconomic variables. This will give you a better idea of how macroeconomists think about the way the world works.

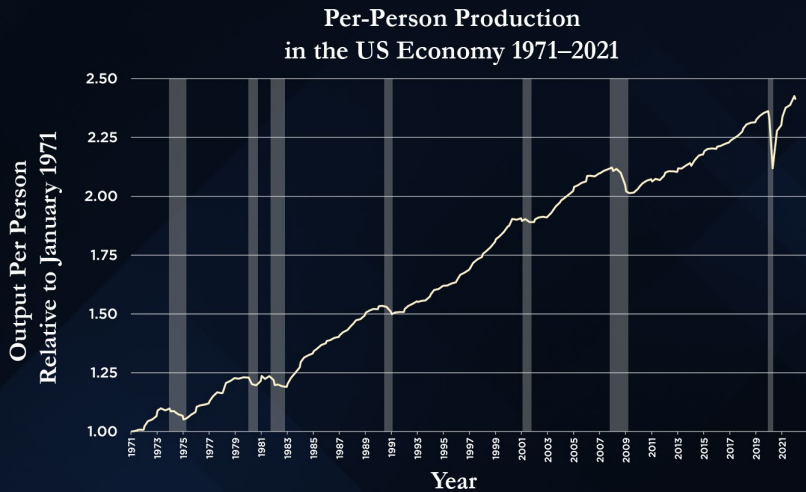


# What Is Macroeconomics?

Macroeconomics provides economists with a new lens that brings clarity to the world. For example, over the last four decades, China has increased the per-person production of its economy by a factor of 10. What economic reforms did China take that enabled it to become an economic superpower so quickly? In 2022, inflation climbed to levels not seen since the early 1980s, and stories about rising inflation dominated newspaper headlines. What changes in the economy led to these concerns about rising inflation, particularly after almost a decade of worrying that inflation was too low? These are questions that you can answer more clearly by improving your knowledge of macroeconomics.

Economics as a whole offers two different lenses to examine the world with. *Macroeconomics* is a relatively new word, dating from 1945, but it encompasses older topics, such as money and business cycles, that had already been studied for many decades. Studying macroeconomics leads students to ask significant questions. The counterpart to macroeconomics is microeconomics. Microeconomics looks at the production decisions of individual firms, while macroeconomics looks at the production decisions of an entire economy.

Consider how production in the US economy changed from 1971 to 2021. The first thing to note is that production per person in the economy has grown steadily over time. Since 1971, per-person output has increased by a factor of around 2.35 over 50 years. There is nothing automatic about such steady growth. For one thing, not all countries prosper over time. For example, Nigeria was growing faster than the United States in the 1970s but then began a long period of declining output per person. At one point in the mid-1990s, per-person output in Nigeria was only about 15% of what it had been at the start of the 1970s. Thus, even when the economy did eventually rebound in the 2000s, Nigeria had still not achieved the growth path of the United States. Zimbabwe was on a path similar to the United States until 1996. However, from 1997, output per person fell as sharply as happened for Nigeria. Since Zimbabwe's severe downturn is more recent and the recovery has been weaker, Zimbabwe's economy was producing marginally less output per person in 2020 than it produced in 1971.



Other countries have achieved sustained and dramatic economic growth over these 50 years. For example, South Korea grew extremely fast over this entire period. This sustained and rapid long-run growth means that South Korea's economy produces 16 times as much per person as it did 50 years ago. This works out to a per-person growth rate of 6% a year. Now look at China, where output per person in 2020 reached around 10 times what it was in 1971. This is about four times more transformative than the United States over the same period. Moreover, China's situation did not begin to change until the end of the 1970s, almost a decade later than South Korea. That was when China embarked on a new rapid growth path beginning with the economic reforms of Deng Xiaoping in 1978. India's transformation started even later than China's, beginning in the 1990s. Two decades of sustained economic growth resulted in India's production per person being four times what it was 50 years earlier.

Macroeconomists treat questions related to economic growth as being of paramount importance for human welfare. There has never been a period in human history in which so many people have been raised out of abject poverty in such a short period of time as what has happened over the past 50

years in East Asian countries such as South Korea and China. If that success can be replicated over the next 50 years in South Asian and sub-Saharan African nations, there is the opportunity to save billions more from abject poverty. That said, sometimes solving one vexing problem can lead to the creation of a new one. Suppose a global middle class of 5 billion people comes into existence. The cumulative impact of so many people behaving more like people in wealthy countries will also pose new challenges for the resources and carrying capacity of planet Earth as a whole. Clearly, growth pathways will also need to become more sustainable.

Consider how production per person evolved in the United States over the past 50 years, noting the up-and-down fluctuations in production. The overall trend is rising, but there are also drops. The most noticeable downturns are officially labeled as recessions. Recessions vary considerably in severity, whether that is measured in terms of the length of the recession or the magnitude of the downturn. The 1990/1991 recession and the 2001 recession were among the shortest recessions in US history, each lasting about eight months. Both also resulted in relatively small drops in output. At other times, however, recessions can be long and devastating. The oil shock recession of the 1970s, the double-dip recessions of the 1980s, and the 2008/2009 global financial crisis all lasted more than a year and resulted in about a 4% drop in economic activity.

The two most recent downturns are particularly noticeable. The year 2008 saw a sharp and long-lasting drop in real gross domestic product (GDP) that took almost six years to recover from. The National Bureau of Economic Research (NBER) Business Cycle Dating Committee officially declared the COVID-19 recession as covering the period from February 2020 to April 2020, the shortest US recession on record. The economy fully recovered to its previous level within a year. However, it was still the most severe drop since the end of World War II when measured in terms of the decline of real GDP from peak to trough.

## Studying Economic Fluctuations

The second main area of macroeconomics is the study of economic fluctuations. Although production in the United States has grown steadily over time, occasionally, the economy stops growing and even shrinks before eventually growing again. Economists refer to the overall pattern of reversal

## 1. Viewing the World through a New Lens

and rejuvenation as a business cycle. The peak represents the point at which economic activity begins to decline. The trough represents the point at which the decline in economic activity begins to reverse and the economy starts increasing in size again. The time period between the peak and the trough is what economists officially label as a recession. The period between the trough and the next peak is an expansion. The early period of the expansion (where real GDP is still trying to reach the pre-recession heights) is the recovery.

In the United States, the official designator of business cycles is the NBER's Business Cycle Dating Committee. Economists on this committee identify key turning points in overall economic activity to designate as the peaks and troughs. The committee has identified 68 business cycle turning points during its 126-year history between 1894 and 2020. From 1970 to 2020, the United States had seven recessions and seven recoveries. A rule of thumb says that the decline in economic activity must last for two quarters—a half year—before the decline is considered a recession. However, this is not always the case, as you saw with the 2020 recession, which the NBER deemed to have lasted only two months, from late February to early April. In that case, the decline in economic activity was so noticeable that there was not much doubt about whether that was a recession.

Recession		Duration of Recession (months from peak to trough)	Decline in Output (percent from peak to trough)	Duration of Next Expansion (months from trough to peak)
Peak	Trough			
Nov. 1973	Mar. 1975	16	4.9	58
Jan. 1980	Jul. 1980	6	2.5	12
Jul. 1981	Nov. 1982	16	3.0	92
Jul. 1990	Mar. 1991	8	1.4	120
Mar. 2001	Nov. 2001	8	0.0	72
Dec. 2007	June 2009	18	4.1	128
Feb. 2020	Apr. 2020	2	10.6	26*

\* As of June 2022

No two business cycles are alike in terms of what caused the downturn, how long it lasts, or how strong the eventual picking up of economic activity is. However, by understanding the causes of a recession, macroeconomists can help identify policy and institutional changes that can make the economy more robust to a future shock of the same type. In some cases, the cause of a recession is clear: The 1973 recession was caused by a sharp rise in oil prices stemming from an embargo on oil exports by the Organization of the Petroleum Exporting Countries. The 2001 recession was caused by a sharp fall in the stock prices of internet companies that had enjoyed an extremely rapid rise in the late 1990s. The fall was compounded by the attacks of September 11. The 2008/2009 recession was caused by a collapse in housing prices together with the related fallout for the highly leveraged financial sector. The 2020 recession was caused by the COVID-19 pandemic.

In other cases, the causes of recession are more complex. In the early 1980s, the recession was initially caused by higher oil prices stemming from the disruptions caused by the Iranian Revolution of 1979. That later combined with another recession caused by aggressive policy actions from the Federal Reserve raising interest rates to bring down the rate of inflation in the economy. Economists call this a double-dip recession—one where the recovery from the first recession is not able to establish a firm footing before the second slump begins.

## Macroeconomic Variables

In trying to understand what drives economic growth and economic fluctuations, macroeconomists also focus on the connections between macroeconomic variables: how changes in one aspect of the economy impact other areas of the economy. For example, consider the important relationship between the state of the economy and unemployment. During the 18 months of the 2008/2009 recession, the unemployment rate increased from 4.9% to 9.9%; that is, the number of unemployed workers across the country doubled, from 7.6 million to 15.2 million. An even larger, though more short-term, increase in unemployment occurred in the US economy in the early part of 2020. The unemployment rate climbed from 3.5% to 14.7% in two months. That translated into almost 18 million workers becoming unemployed between mid-February and mid-April of 2020.



Another important relationship is between the state of the economy and the inflation rate—a measure of how fast prices of goods and services rose from one year to the next. Inflation rose before the recessions of 1974 and 1979 and during the recession of 1991. By contrast, preceding the more recent recessions of 2001, 2008/2009, and 2020, inflation did not noticeably rise. However, inflation fell during all six recessions. Notice also that the US economy experienced a long-term trend of falling inflation (or disinflation): from the high inflation rates of 6% to 10% in the 1970s to around 4% in the 1980s and down to around 2% by the 2010s. Falling prices during the 2008/2009 recession were particularly noteworthy because the economy experienced negative rates of inflation—deflation.

Another macroeconomic variable that economists pay attention to is the interest rate. When interest rates rise, for example, it is more expensive to borrow funds to buy a house or a car. Thus, many people may postpone such purchases. Economists pay particular attention to the federal funds rate: the interest rate banks charge each other on short-term loans. The ups and downs in the federal funds rate are closely correlated with the ups and downs in the economy. First, rates typically rise before each recession and then decline

during and after each recession. Second, as with the inflation rate, there are long-term trends in the interest rate. The interest rate rose in the 1970s and early 1980s, but beginning in the mid-1980s, it began a downward trend.

During the 2008/2009 recession, the federal funds rate fell to close to 0%, which is the lowest it can go. It remained there for around five years before rising gradually, albeit to levels that were still low by historical standards. As soon as the COVID-19 pandemic hit the economy in the first quarter of 2020, the federal funds rate fell back again to its lower bound of zero.

# Intercountry Connections

A key variable in how countries interact with one another is the exchange rate—the price of one country’s currency (e.g., dollars) in terms of another country’s currency (e.g., yen). From 1945 to 1971, the major economies of the world operated under the Bretton Woods Agreement. This was a system whereby 43 countries agreed to fix the value of their currencies to the US dollar, which in turn was convertible into gold at the fixed rate of \$35 per ounce. In 1973, the Bretton Woods system collapsed entirely. Countries went in many different directions with their exchange rates.

Some, such as the United States, chose to allow their currency’s value to be determined freely in a market by supply and demand conditions. This led to the creation of the enormous foreign exchange market, where trillions of dollars’ worth of currency are traded each day, dwarfing the annual production of the global economy. Other countries—particularly developing economies—chose to fix their currency’s value to a major currency such as the dollar, the pound, or (at the time) the French franc. Yet others, especially European nations, experimented with setting up a smaller Bretton-Woods-like system among themselves in the 1980s—before that, too, collapsed in the early 1990s. They regrouped and, in 1999, launched the most ambitious international finance experiment of the new millennium: the creation of the eurozone, whereby 19 countries chose to give up their own currencies and adopt a common currency, the euro.

# Causal Relationships

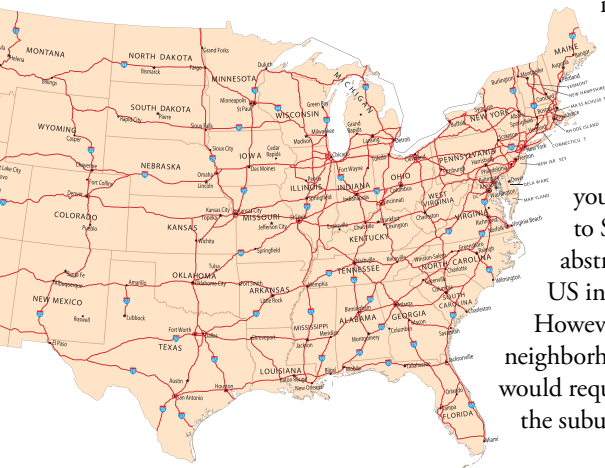
Macroeconomists also look beyond the interconnections of macroeconomic variables to find causal relationships: How does a change in one variable (*a*) impact another variable (*b*), all else equal? Economists love to remind people that correlation does not imply causation—just because two variables move together does not mean that one is driving the behavior of the other. In macroeconomics, controlled experiments are hard to come by. What you can do instead is take advantage of a natural experiment: a situation where a policy could arbitrarily impact some groups but not others—creating a synthetic treatment group and control group to compare and contrast. However, even natural experiments are hard to find in macroeconomics. Even if different states within the United States adopt different policies on how much to tax capital, you can't learn much about the effectiveness of the policy if the state that adopted the lower capital tax also differed in many other ways from the one that did not.

As a result, macroeconomists often have to use economic models to understand the channels through which a particular policy could impact the economy and to what degree. A good analogy compares economic models to maps. Models, like maps, are abstractions of a more complex reality and come in all shapes and sizes: They can be complex or simple; they can highlight many different features of the area being covered or focus on a single feature. Every model can be useful

in its own right but may be useless for a different task.

Furthermore, the usefulness of the model does not necessarily depend on its complexity. For example, if you want to drive from Boston to San Francisco, you need an abstract map that lays out the US interstate highway system.

However, if you are visiting different neighborhoods in your community, you would require a detailed map of the suburbs.



## READING

- Blinder, Alan. "What Did We Learn from the Financial Crisis, the Great Recession, and the Pathetic Recovery?" *The Journal of Economic Education* 46, no. 2 (2015): 135–149.
- Heilbroner, Robert L. *The Worldly Philosophers: The Lives, Times and Ideas of the Great Economic Thinkers*. Simon and Schuster, 2011.
- Ouliaris, Sam. "Economic Models: Simulations of Reality." *Finance & Development*. International Monetary Fund, June 2011.
- Rodrigo, Chris. "Micro and Macro: The Economic Divide." *Finance & Development*. International Monetary Fund, May 2018.
- Spencer, Roger W., and David A. Macpherson, eds. *Lives of the Laureates: Thirty-Two Nobel Economists*. MIT Press, 2020.
- Stiglitz, Joseph E. "Some Lessons from the East Asian Miracle." *The World Bank Research Observer* 11, no. 2 (1996): 151–177.

# 2

## GDP: THE ECONOMY IN A SINGLE NUMBER?

**T**he intellectual origins of GDP come from the attempts to understand the macroeconomic devastation caused by the Great Depression in the 1930s. Economist Simon Kuznets devised a way to quantify the damage caused by the Great Depression in a single number. Today, this measure, albeit honed and improved, is widely used by economists as the best measure of economic activity. It is also the key data input used to understand many of the macroeconomic phenomena related to economic growth and economic fluctuations that you will study in this course.



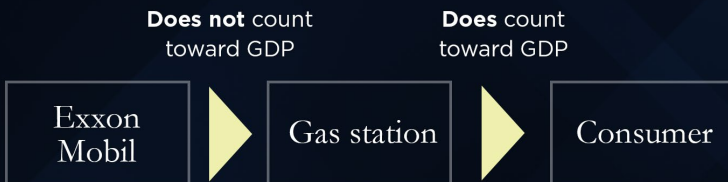
## What Is GDP?

The technical definition of GDP is the value of all newly produced goods and services in the domestic economy during a specific time period. GDP measures the value of the newly produced goods and services in the economy. It is a flow measure: a measure of how many new goods and services were added to the economy during the specified period. Moreover, GDP is a measure of what is produced in the domestic economy only. For example, if Ford, an American automobile company, produces cars in Mexico or Canada, those do not count toward US GDP. However, if Toyota, a Japanese company, builds cars in Ohio, those do count toward US GDP.

Until the early 1990s, newspapers used to talk about a measure called gross national product (GNP). This is a measure of output produced by domestic entities, regardless of where the production took place. When the United States moved from using GNP to using GDP in the early 1990s, complicated products such as automobiles were increasingly assembled using components from more than one country, and GDP better captured where the economic activity took place.

## Real GDP

To calculate GDP, economists add up spending on all final goods and services produced in the domestic economy using market prices. A final good or service is one that, during the period in concern, will neither be used as an intermediate input for producing another good or service nor resold. The focus on final goods is to avoid double counting. For example, if a gas station



## 2. GDP: The Economy in a Single Number?

buys gasoline from Exxon Mobil and you buy that same gasoline for your car in the same period, the gas station's purchase of gasoline is not counted because it is not a final good (the gas was resold to you).

Because GDP only measures market transactions, it omits nonmarket contributions to the economy. The Bureau of Economic Analysis (BEA) estimates that nonmarket household production, including taking care of family members, reached \$1 trillion by 1979 and more than \$4 trillion by 2013. That's an omission equal to almost 25% of stated GDP. There's also the shadow economy, where people do get paid for production but don't report it. Economists try to correct for these omissions where possible. For example, in the United States, the imputed rental income of people who live in their own homes is incorporated into the calculation of GDP to avoid a situation where renters contribute to GDP but homeowners do not.

Moreover, because GDP measures output using the market prices of goods, changes in the price of goods will affect the value of GDP even if there was no underlying change in the quantity of goods produced. The measure of output that corrects for price changes is real GDP, sometimes called constant-price GDP. In fact, it's real GDP that appears most often in news reports. To avoid the confounding effect of price changes, real GDP uses a single set of prices from a reference year to calculate the dollar value of goods and services produced in the economy in different periods. The reference year, or base year, is arbitrarily chosen.

In practice, using a base year price to calculate real GDP becomes complicated because of the introduction of new goods, the obsolescence of some existing products, and quality changes of continuing products. For example, suppose you choose the base year to be 1994 in calculating real GDP for 2021. What would you use as the base-year price of a 2021 iPhone? An economist in 1994 might have said such a device could be worth a billion dollars. However, in 2021, it was available for a few hundred dollars. You can try to avoid this by choosing a more recent year to be the base year. However, you will then have problems calculating real GDP for 1994. What would be the 2021 price of a formerly state-of-the-art Sony Walkman cassette player?

Quality improvements are also a problem. An Apple computer you bought in 1991 would cost you \$2,000, just as the Apple computer you bought in 2021 did. However, the 2021 version is unfathomably more powerful and capable

## 2. GDP: The Economy in a Single Number?

compared to the 1991 computer. Thus, its contribution to real GDP may stay the same unless statisticians can capture the fact that the production of computing power is vastly improving.

# Calculating GDP

Economists have come up with techniques that effectively minimize the impact of these quality improvements and new products on the accuracy of real GDP. For example, because many companies and individuals now use cloud services to store their data on remote sites rather than on their own machines, economists have to account for the fact that consumers are buying a service (more storage space on a remote computer) rather than a good (a larger hard drive) for the same task. They also need to make sure that cloud computing and advances in the speed at which customers can access data stored on the cloud are captured in real GDP calculations.

In calculating GDP for the economy, the BEA divides spending in the economy into four categories: consumption ( $C$ ), investment ( $I$ ), government purchases ( $G$ ), and net exports ( $NX$ ), where net exports equal exports minus imports ( $NX = X - M$ ). This leads to a famous macroeconomic identity:  $GDP = C + I + G + NX$ .

Quite often, a news release about a sharp rise or fall in GDP will contain supplementary information about which components of GDP were responsible for the rise or fall. *Consumption* refers to the spending on goods and services by households—it is the largest component of GDP. It has risen from being about 60% of GDP in 1960 to around two-thirds of GDP in 2020. Consumption includes purchases of durable goods that provide consumption services over time (TVs, cars, or refrigerators), nondurable goods (food, beverages, or gasoline), and services (haircuts or medical care).

*Investment* refers to the amount of new capital (machines, equipment, raw materials, etc.) that firms add to the existing capital stock. It is the most volatile component of GDP. In times of economic weakness, investment typically falls, as it did in the recessions of 2008/2009 and 2020. In times of economic strength and recovery from recessions, investment typically booms. There are a couple of idiosyncrasies about investment to note. For instance, spending by individuals on housing is classified as investment rather than consumption.

## 2. GDP: The Economy in a Single Number?

In other words, a new house or apartment is thought of as a capital good that produces housing services in the future rather than as a consumption good. In fact, in the period surrounding the 2008/2009 recession, investment increased when the housing market boomed but experienced a dramatic fall after the real estate crash led to the 2008/2009 recession.

Conversely, private spending on education (about \$300 billion in the United States) is not considered to be investment. Instead, it is categorized as consumption of a service. Moreover, investment here does not refer to financial investment (i.e., buying stocks or bonds or putting money into retirement accounts). Only the purchases of goods and services count toward GDP.

*Government purchases* refers to spending by federal, state, and local governments on new goods and services. These include military spending on personnel, aircraft, and weapons; the building of highways, roads, and bridges by state governments; and spending on education and police forces by local governments.

Government purchases as a share of GDP have been trending downward over the past 50 years—from around 24% of GDP to around 18% of GDP. Moreover, when the government share of the economy suddenly increases in peacetime, it's mostly not because the government has declared all-out war on the downturn via new contributions to GDP. For example, you might imagine the 2020 increase reflects the trillions of dollars the government spent on pandemic relief checks and unemployment benefits to offset the economic harm imposed by the shutdowns in 2020.

Interestingly, however, GDP does not include the value of transfer payments, such as unemployment insurance. This is because no new goods are produced. The same applies to taxes, which are merely a transfer payment from you to the government. The reason why the share of government purchases in GDP went up was simply because government purchases remained relatively unchanged when other components of GDP dropped sharply. Interestingly, Kuznets argued that government spending should not be included in GDP. However, by 1940, war-related production was becoming an economic activity of a scale that had to be incorporated if GDP were to remain an accurate

## 2. GDP: The Economy in a Single Number?

calculation of economic activity. During World War II, government purchases accounted for more than 40% of GDP, but the percentage since then has never risen above 25%.

The fourth and final component of GDP adds the net contribution of exports minus imports, or net exports. Economists want the total production to account for goods and services produced in the United States even when they are sold abroad. Conversely, they subtract imports because those are goods and services produced abroad. For example, suppose you buy a bicycle from a shop that imports the frame from abroad but buys the tires in the United States. The data collectors can observe how much you paid for the bicycle and count that as consumption. However, to accurately measure GDP, they would need to subtract off the imported bike frames because that part of the bicycle was not produced in the United States. Net exports can be negative if imports exceed exports. In fact, net exports for the United States have been consistently negative every single year since the price of imported oil moved sharply higher in the 1970s.

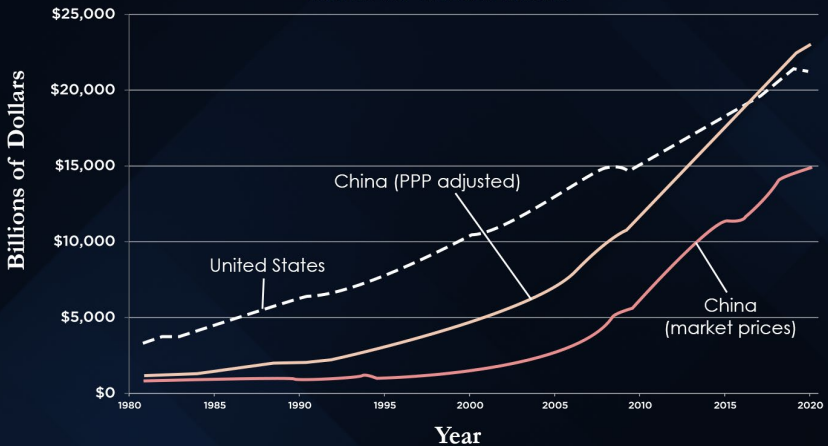
## Comparing Economies

Suppose you want to compare GDP for the economies of Sri Lanka and the United States. You can look up data for GDP for the United States (in dollars) and for Sri Lanka (in rupees). You can use the exchange rate between dollars and rupees to express both numbers in the same currency—say, dollars. However, this is still problematic for two reasons. First, the market exchange rate is itself a price of one currency in terms of the other. Thus, once again, you are allowing changes in prices to affect the measure of quantity. For example, during the pandemic, the value of the Sri Lankan currency fell from 150 rupees per dollar to 200 rupees per dollar. Even if the quantity of goods and services produced in Sri Lanka had not changed during the pandemic, you would get a smaller dollar value for Sri Lankan GDP from the fact that you are dividing by 200 rupees per dollar rather than 150 rupees per dollar.

Second, there are substantial price differences (especially in services) across countries. For example, a haircut may cost \$20 in Wellesley, Massachusetts. However, in Sri Lanka, that same haircut may only cost the equivalent of \$5. If you allow the same service (a haircut) to contribute much more to GDP in some countries and much less to GDP in other countries, you will not be able

## 2. GDP: The Economy in a Single Number?

### Real GDP Comparison: China vs. United States



to compare accurately across countries. Put another way, Sri Lanka's GDP may not be as low if you consider the fact that goods and services typically cost less there.

What is the solution? Where prices are changing from country to country, you can use one set of prices (base country prices) to value goods produced in different countries. If you use the prices of goods taken from the United States to value goods and services produced in Sri Lanka, you have a more accurate comparison. This type of calculation is called GDP at purchasing power parity (PPP). In 2017, the International Monetary Fund (IMF) reported in their World Economic Outlook that, for the first time, China's GDP measured at PPP (\$19.8 trillion) exceeded the US GDP measured at PPP (\$19.5 trillion). At market prices, however, China's GDP was only about 60% as large as the US GDP (\$12.2 trillion). This shows the importance of using a measure that lends itself to a more accurate comparison across countries.

Even after you correct for differences in purchasing power across countries, you need to account for the fact that the population of China (1.4 billion people) is over four times as large as the population of the United States (325

## 2. GDP: The Economy in a Single Number?

million people). To account for this, economists use GDP per capita. This is calculated simply as GDP divided by population. You would do the same for real GDP (real GDP per capita) to compare living standards within a country over time and for GDP at PPP (per capita GDP at PPP) to compare living standards across countries at a point in time. The per capita GDP at PPP for the United States would be \$19.5 trillion divided by 325 million, or \$60,000. Meanwhile, the per capita GDP at PPP for China would be \$19.7 trillion divided by 1.4 billion, or \$14,000. Thus, per capita GDP in China is only about one-fourth of per capita GDP in the United States.

## Limitations of GDP

Real GDP per capita is used by economists as a proxy for how much an economy produces per person over time. A purchasing-power-adjusted measure of real GDP per capita would be used to compare across countries. However, there are still limitations. For one thing, GDP per capita is a simple average that masks underlying inequalities. An economy where 1 person out of 10 earns \$1 million while the remaining 9 earn \$1,000 will have a misleadingly high per capita GDP of \$100,900. In both China and the United States, the output of the economy is not equally divided among the population in practice—and division inside many countries has become less equal in recent decades. Thus, macroeconomists also look beyond GDP per capita at other measures of economic inequality.

Also note that a high GDP per capita is not synonymous with a high quality of life. In theory, a country with high GDP per capita should be more able to deliver beneficial things than a country with low GDP per capita. However, in practice, that is not always the case. China, for example, achieved a remarkable increase in real GDP by transforming itself from an agrarian, state-controlled economy in 1978 into the hub of manufacturing goods for the world economy in later decades. GDP records the gains that come from building more manufacturing plants or mining copper or harvesting timber. However, it does not capture the smog, water pollution, or deforestation costs that may accompany the economic activity.

Economists do calculate another measure called net domestic product. However, this adjustment has always been focused only on subtracting out the wear and tear on capital goods used to produce output. National production

## 2. GDP: The Economy in a Single Number?

accounts have not typically included an assessment of wear and tear on nature or people. An international effort to focus on quality-of-life factors began in 1990 with the United Nations' Human Development Index. Moreover, in 2012, the United Nations began issuing reports based on the Inclusive Wealth Index that attempts to include the stocks of natural resources and human capital. These measurements are much trickier since they often involve assigning a value for something that was not bought or sold in a market, such as an uncut tree in a forest. However, even the developers of these alternative measures tend to see them as adjustments, or complements, to GDP.

### READING

Bureau of Economic Analysis and US Department of Commerce.  
“Measuring the Economy: A Primer on GDP and the National Income and Product Accounts.” December 2015. [https://www.bea.gov/sites/default/files/methodologies/nipa\\_primer.pdf](https://www.bea.gov/sites/default/files/methodologies/nipa_primer.pdf).

Coyle, Diane. *GDP: A Brief but Affectionate History*. Princeton University Press, 2015.

Jackson, Steven F. “How to Measure a ‘Giant?’” *Education about ASIA* 25, no. 3 (Winter 2020).

McCarthy, Julie. “The Birthplace of ‘Gross National Happiness’ Is Growing a Bit Cynical.” National Public Radio, February 12, 2018. <https://www.npr.org/sections/parallels/2018/02/12/584481047/the-birthplace-of-gross-national-happiness-is-growing-a-bit-cynical>.

OECD. “Greening Productivity Measurement: Environmentally Adjusted Multifactor Productivity Growth.” September 2016. <https://www.oecd.org/environment/indicators-modelling-outlooks/EAMFP%20brochure%20WEB.pdf>.

# 3

## PRICES: HOW TO QUANTIFY THE DANGERS OF INFLATION

**T**his lecture will first discuss how to summarize the prices of all goods and services in the economy into a single number called the price level of the economy, and then how to calculate inflation using the price level. You will also learn about how a price index is constructed, the different price indices that economists use to calculate the rate of inflation, and why economists worry about the inflation rate being too high or too low in an economy.

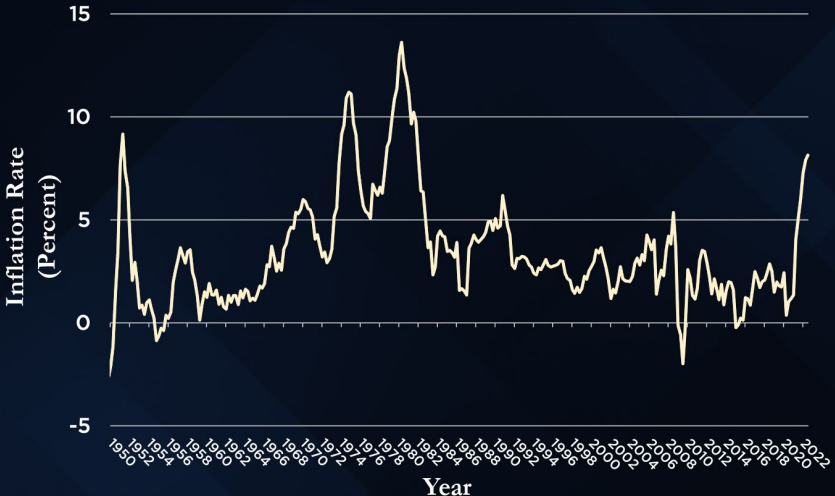


## What Is Inflation?

In October 1978, President Jimmy Carter addressed the American public on television, saying:

Good evening. I want to have a frank talk with you tonight about our most serious domestic problem. That problem is inflation. Inflation can threaten all the economic gains we've made, and it can stand in the way of what we want to achieve in the future. This has been a long-time threat. For the last 10 years, the annual inflation rate in the United States has averaged six and a half percent. And during the three years before my inauguration, it had increased to an average of eight percent.

Inflation in the United States since WWII





The problem of inflation got worse before it got better, rising to 14.8% in March 1980, the highest level reached since just after World War II. In contrast, a generation later, in November 2008, the problem had entirely changed. A front-page story in *The Wall Street Journal* began: “US consumer prices in October registered their largest single-month decline since before World War II, a broad-based drop that creates a new challenge for policymakers: avoiding a bout of deflation, a prolonged period of falling prices throughout the economy.”

One generation’s worries that prices were rising too fast had been replaced by a new generation’s worries that prices were falling. These two stories illustrate the importance of knowing more about inflation. Inflation is simply the rate at which prices are rising in the economy. Think of it like your blood pressure: Too high is not good, too low is not good, but an abnormal reading is usually a sign of some other underlying imbalance in the system.

## Calculating the Consumer Price Index

To calculate inflation, you need an accurate measure of the price level of an economy. The most commonly used measure is the consumer price index (CPI). In the United States, the CPI is calculated every month by the Bureau of Labor Statistics (BLS). It is a measure of the prices faced by a typical consumer for a “basket” of goods and services. The word *index* tells you that the CPI gives you the price of the basket compared to its price in the base year rather than the actual price of the basket or the average price of an item in the basket. Thus, the CPI for any current period is the cost of the CPI basket of goods for the current period divided by the cost of that same basket for a base year. The result is multiplied by 100 so that the CPI is calibrated to always be 100 in the base year. Currently, the base year for the CPI is 1983. Thus, a value for the CPI of 150 in 1994 and 200 in 2006 would tell you that the price level in 1994 was 50% higher than in 1983 and that the price level in 2006 was twice as high as in 1983.

The BLS has carried out an extensive study of consumer purchases since 1913. In every quarter of a two-year reference period, it surveys a large sample of 24,000 consumers about their spending habits. Moreover, it asks another 12,000 consumers to keep a detailed track of their purchases over two-week intervals in a spending diary. The spending habits gleaned from these sources are used to construct a collection of goods and services that form the reference CPI basket. Each item’s relative importance (weight) in the basket reflects consumer purchasing habits. For example, both salt and chicken will be in the CPI basket. However, chicken will be weighted more heavily since consumers buy more pounds of chicken than pounds of salt. Once the reference-year basket is formed, it stays the same until the next revision occurs, which can be more than a decade. After the CPI was set up in 1919, revisions were done in 1940, 1953, 1964, 1978, 1987, 1998, and 2018.

The BLS then collects data on the prices of the individual goods and services in the CPI basket from stores and updates the index every month. The stores are chosen to be representative of the places where consumers shop for goods and services. The BLS uses telephone surveys to identify the establishments. For administrative ease, the stores are located in urban and metropolitan areas that cover about 93% of the US population. Thus, the CPI is sometimes labeled more specifically as the CPI-U (the CPI for all urban consumers). The

### 3. Prices: How to Quantify the Dangers of Inflation

list of individual goods and services in the CPI basket includes thousands of items. Thus, the BLS groups them into eight broad categories—housing and utilities take up the largest share, followed by transportation and fuel, and then food and beverages. When all three of these categories show a sharp rise in prices, as happened in 2021 and 2022, the overall CPI increases sharply as well.

The CPI can also be used to make more accurate comparisons of monetary amounts over time. For example, say there's a student whose parents graduated from college in 1985. The parents tell her that their tuition was \$10,000 a year, compared to the \$50,000 a year they have to pay for her tuition in 2020. Is it five times more expensive for that student to go to college than it was for her parents? A meaningful answer requires that you account for the fact that a dollar in 2020 does not buy as much as a dollar did in 1985.

Here's another example: The federal minimum wage was \$1.60 per hour in 1970, whereas in 2020, it was \$7.25 per hour. Is a minimum wage worker in 2020 getting paid four to five times more than a minimum wage worker was in 1970? Again, consider the fact that a dollar in 2020 does not buy as much as a dollar did in 1970. Economists describe this type of transformation as making a real comparison rather than a nominal comparison. The CPI helps you make this adjustment, allowing you to correct for inflation in the economy.

Consider the college tuition example: The CPI for 1985 was 106.3, and the CPI for 2020 was 255.1. You can then find the 2020 equivalent of \$10,000 in 1985 as being 10,000 divided by the CPI in 1985 and multiplied by the CPI in 2020. This is approximately \$24,000. Thus, the real cost of college tuition has basically doubled (from \$24,000 to \$50,000) instead of the fivefold increase (from \$10,000 to \$50,000) that the nominal cost comparison revealed.

In the minimum wage example, the CPI value for 1970 was 38.6, and the CPI for 2020 was 255.1. You can then find the 2020 equivalent of \$1.60 in 1970 by dividing by the CPI in 1970 and multiplying by the CPI in 2020. This is around \$10.60. This shows that a minimum wage worker in 1970 earned a higher real wage than did a minimum wage worker earning \$7.25 in 2020. Thus, the real price of college doubled, while the real value of minimum wage jobs declined by more than 30%.

### 3. Prices: How to Quantify the Dangers of Inflation

Sometimes, firms and governments can do the CPI calculation in reverse and use the CPI to adjust a stream of payments over time to keep the real value of the payments constant. For example, a teacher may retire at the age of 60 and be entitled to a pension of \$50,000 a year in retirement. Suppose that the government wants to make sure that the purchasing power of this pension does not erode over time. If the CPI in the year of retirement was 220 and the CPI five years after retirement was 250, then the pension in that year should be scaled up to be \$50,000 divided by the retirement year CPI and multiplied by the CPI in five years. That number would be \$56,800. The CPI is used in this manner to index Social Security payments, public sector pensions, and payments on treasury inflation-protected securities.

## Calculating Inflation

Suppose you have new CPI data that was released for July 2022. Inflation over the past 12 months is given by the price level change over the last 12 months divided by the price level at the start of the 12 months. Sometimes, economists want to know how much inflation there was in the most recent month. They simply take the price level change over the last month divided by the price level in the previous month. Note that this is a monthly change rather than an annual change, meaning the magnitudes are different. To avoid confusion, sometimes news stories will express the monthly change as an annualized rate, which is a fancy way of saying, “By how much would prices rise over a year if they increased for the next 11 months at the rate that they increased over the last month?” This is roughly similar to multiplying the monthly number by 12 for low inflation rates. However, due to the effect of compounding, the result of multiplying by 12 becomes less accurate as the monthly inflation rate increases.

There are several other measures of prices that economists construct and use. The first is the personal consumption expenditures (PCE) deflator (sometimes called the PCE index). This has become the Federal Reserve’s preferred method of measuring the price level. The PCE deflator is the consumption portion of nominal GDP divided by the consumption portion of real GDP. Recall that nominal GDP values goods and services produced in the current period at current prices, while real GDP values those same goods and services at base-year prices. The idea behind the PCE is similar to the CPI: It

### 3. Prices: How to Quantify the Dangers of Inflation

compares the cost of a “basket” that includes all goods and services consumed in the domestic economy in the current period to the cost of that basket in the base year. However, the PCE deflator has one important advantage.

In calculating the CPI, the BLS uses a basket of goods and services that captures what the typical consumer would purchase in the reference year. Once the basket is constructed, its contents are fixed. The BLS only needs to see how the cost of that basket is changing over time relative to its base year cost. In contrast, when the BEA constructs the PCE deflator, it constructs the basket for the current period (i.e., what consumers purchased this period) and compares what that new basket would cost today to its cost in the base year. This continual updating of the PCE basket means that the PCE is a better reflection of what consumers are buying now. The CPI basket was a better reflection of what consumers bought in the reference year. This helps the PCE overcome a common criticism that the CPI overestimates the true level of inflation in the economy because it ignores the fact that consumers can change what they consume when prices rise.

Suppose consumers in the reference year bought 50 pounds of chicken and 25 pounds of pork. The CPI basket will contain twice as much chicken as it does pork. Suppose that the price of chicken rises in some subsequent year, but the price of pork does not change. Consumers will switch some of their chicken consumption to pork consumption, say, to 40 pounds of chicken and 40 pounds of pork. The CPI basket (which reflects the reference year spending patterns) will not capture that change. It will still think that consumers buy twice as much chicken as they do pork. The PCE, however, will accurately capture that chicken and pork consumption were equal and should be weighted the same in the current basket. Thus, the CPI will indicate a higher price level than the economy is experiencing because it does not capture how higher prices for a good can reduce consumption of that good. This is why the PCE is considered to be the Federal Reserve’s favorite price index.

Why do economists not use the PCE for all purposes? Because it takes more time to get accurate data on what people bought in the current period. For instance, GDP data for the first quarter, covering January through March, is not released until the end of April and then revised again in May and June. These quarterly numbers are also subject to an annual revision in the middle of the next year. At

### 3. Prices: How to Quantify the Dangers of Inflation

best, if you used the PCE, you would not know how much prices rose in January until three to six months later. By contrast, the CPI is available every month, which is why it remains the overall favored measure of prices.

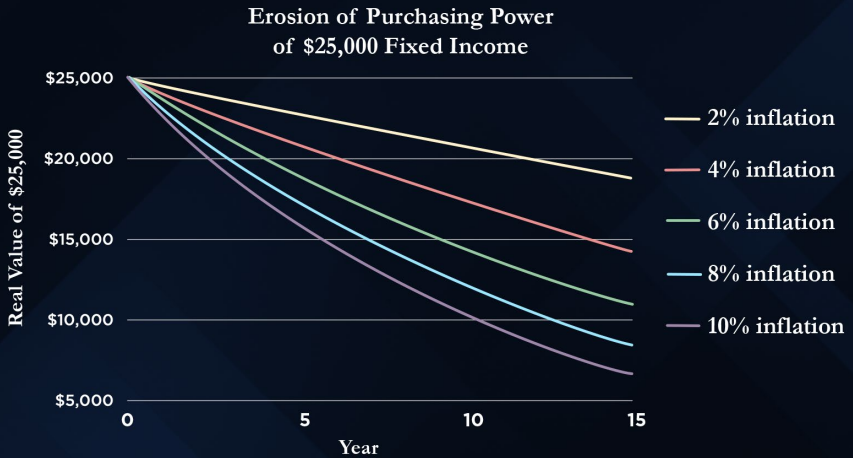
A third measure is the producer price index (PPI). The PPI is similar to the CPI. However, the PPI uses a basket of goods and services that a typical producer would purchase in constructing the index. This means that warehouse costs, shipping costs, and the cost of raw materials are likely to feature much more prominently in determining the value of the PPI. The PPI provides an overview into the pressures that producers face. It can often serve as a harbinger for changes to come in the CPI. When trade barriers or supply shocks affect the cost of raw materials, the PPI will rise much earlier than the CPI and provide a signal to economists about potential price increases that will arise if firms have to pass on higher costs to consumers.

Finally, the core CPI is basically a CPI that omits the cost of food and energy. In other words, the core CPI basket leaves out the food and energy-related items from the CPI basket. It is used because the prices of food and energy tend to be much more volatile than other components of the CPI. If prices of food or fuel rise sharply in March, you may even see a decline in those prices in April. By taking out these unpredictable fluctuations, the core CPI may have a higher signal-to-noise ratio than the CPI does in a given month.

## Importance of Inflation

There are three broad reasons why inflation is a cause for macroeconomic concern. First, inflation erodes the purchasing power of money. Consider a retiree who receives a fixed pension of \$25,000 and compare what happens to the purchasing power of that pension under different inflation rates, ranging from 2% to 10% over a 15-year period. Even a 2-percentage-point change in inflation can have a dramatic impact. At 2% inflation, which has been the Federal Reserve's inflation target for many years, the real value of \$25,000 in 15 years' time is around \$19,000. However, at 4% inflation, it falls to \$14,000. At the 6% rate of chronic inflation President Carter mentioned, \$25,000 falls to \$10,500 in 15 years. It's often disadvantaged groups who bear the brunt of inflation. Minimum wage workers receive wages that are fixed until politicians change them.

### 3. Prices: How to Quantify the Dangers of Inflation



Second, inflation creates winners and losers in the economy. Note that even in the previous example, where inflation hurt the retirees receiving a fixed income, it would help the groups paying the fixed incomes. Taxpayers, corporations, and pension funds all benefit from the lower real value of the payments they make. Similarly, unexpected inflation helps borrowers on a fixed-interest-rate loan while hurting lenders: If you took out a fixed-rate loan to buy a house at an interest rate of 5% when inflation was around 2%, all else equal, you would benefit if the inflation rate went to 4%. Intuitively, this is because the money you pay back has less purchasing power than the money you borrowed. Also, note that a 3% inflation rate in the economy does not mean that each individual price rose by 3%. The price of some goods would increase by more than others. For example, in the late 2000s, the price of housing fell while the price of gasoline and food rose sharply. This would benefit some groups (home buyers, farmers, oil companies) at the expense of others (consumers, homeowners, people with long commutes).

Third, inflation creates uncertainty and shortens planning horizons. For example, an airline company and the pilot's union have to worry about the impact of inflation on their wage contract because the real cost of the wages and pensions they agree upon in their contract will be impacted by inflation that is higher or lower than what they expected. In countries

### 3. Prices: How to Quantify the Dangers of Inflation

with unpredictable inflation, people are reluctant to enter into long-term relationships. This makes it hard to borrow money over a long period of time to buy a house or to build a factory. People prefer to buy and hold gold or land instead of putting money into a bank or lending it out to entrepreneurs. The lack of long-term investments reduces the growth rate of an economy. In theory, an economy where inflation is always 15% a year may lend itself better to long-term planning than an economy where inflation is sometimes 1% and at other times 6%. Empirically, however, higher inflation and more variable inflation are strongly correlated.

## READING

Baker, Dean. *Getting Prices Right: Debate over the Consumer Price Index*. Routledge, 2016.

Bureau of Labor Statistics. "Consumer Price Index Frequently Asked Questions." Accessed December 12, 2022. <https://www.bls.gov/cpi/questions-and-answers.htm>.

Horsley, Scott. "Think Inflation Is Bad Now? Let's Take a Step Back to the 1970s." National Public Radio, May 29, 2021. <https://www.npr.org/2021/05/29/1001023637/think-inflation-is-bad-now-lets-take-a-step-back-to-the-1970s>.

Iacurci, Greg. "Gold as an Inflation Hedge? History Suggests Otherwise." CNBC, June 8, 2021. <https://www.cnbc.com/2021/06/08/gold-as-an-inflation-hedge-history-suggests-otherwise.html>.

Palley, Thomas. "How to Rewrite Economic History." *The Atlantic*, April 1997. <https://www.theatlantic.com/magazine/archive/1997/04/how-to-rewrite-economic-history/376830/>.

Tanzi, Alex. "In the Lockdown Economy, Old Measures of Inflation Get It Wrong." *Bloomberg News*, May 22, 2020. <https://www.bnnbloomberg.ca/in-the-lockdown-economy-old-measures-of-inflation-get-it-wrong-1.1439948>.

# 4

## UNEMPLOYMENT: WHO WANTS TO WORK?

**T**his lecture discusses key macroeconomic variables and focuses on understanding what the unemployment rate is. The three measures of the labor market you will learn about—the unemployment rate, labor force participation rate, and employment-to-population ratio—paint a picture about how the labor market is faring at a particular point in time. This allows economists to place the current performance in historical context. As with measures of prices and output, developing an understanding of the strengths and limitations of the key macroeconomic measures of the labor market helps build a strong foundation for understanding all of macroeconomics.



## What Is the Unemployment Rate?

The unemployment rate is the macroeconomic variable that is most likely to have the most direct impact on any individual's life. A high unemployment rate can make it difficult for people who don't have a job to find one. It will disrupt the entry of new graduates into labor markets and unexpectedly hasten the retirement decisions of older workers. Even for those who hold on to their jobs, it will increase feelings of job insecurity, reduce mobility, and hold down wages.

In the United States, the unemployment rate is calculated by the BLS in the Department of Labor using data from two surveys that are conducted by the experts at the US Census Bureau of the Department of Commerce. The BLS puts out a report called "The Employment Situation" on the first or second Friday of the month. This report draws on survey data from households and business establishments. The Current Population Survey (CPS) covers a nationally representative sample of about 60,000 households in the United States, containing more than 100,000 individuals. Note that the unemployment rate is not calculated using unemployment benefit recipient data from the states.

For the CPS, every month, data collectors who are experienced in survey methods reach out to sample households and interview them either in person or over the phone. Surveyors ask the household about their labor market activities during a reference week, which is typically the middle of the month they are being surveyed. People's answers to the questions determine how they are classified in terms of their current status in the labor market. Questions may include: "Did you do any work for pay or profit?" or "Did you do any unpaid work in a family business or farm?"

The second survey is the Current Employment Statistics (CES) survey. It covers a sample of around 150,000 businesses across the United States, covering almost 700,000 work locations. The data from this survey are used to provide monthly estimates of employment patterns across various sectors of the economy, hours of employment, and earnings.

## Employment Status

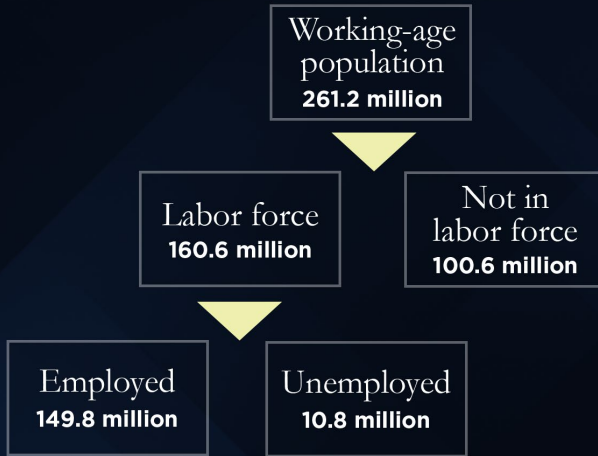
When discussing unemployment or the labor market, the focus is on the civilian noninstitutionalized working-age population. This is the number of individuals 16 years of age and over who are not inmates of institutions and who are not actively serving in the military. The working-age population at the end of 2020 was 261.2 million, compared to a total population of 330.9 million. In other words, in 2020, there were 261.2 million people in the United States who were theoretically able to work for pay. The working-age population can be split into three groups: people who are employed, people who are unemployed, and people who are not in the labor force.

To be counted as employed, a person must have worked for pay or profit during the reference week of the survey. There is no minimum hour requirement to be counted as an employed worker. If you had a job but did not work during the reference week because of atypical circumstances, you would still be counted as employed. There is one exception to the “worked for pay or profit” requirement. Individuals who work in a family-owned business or farm for more than 15 hours a week, such as a high-school-aged child who helps out with livestock on a farm, are considered employed even though they are not paid for their work.

An unemployed person is one who does not have a job but is actively looking for work. People are not considered to be unemployed if they lack a job but are not actively looking for work (e.g., students). During the prior four weeks, you must have taken steps such as filling out a job application, submitting a resume for a position, interviewing for a job, or contacting an employer or employment agency about a job opening. One exception to the requirement for active job search concerns workers who have been placed on temporary unpaid leave from their jobs. This was a particularly common feature during the early days of the pandemic.

The sum of all employed and unemployed people in the working-age population is the labor force. In 2020, the US economy had 149.8 million employed people and 10.8 million unemployed people, for a total labor force of 160.6 million people. Notice that this total labor force is much smaller than the 261.2 million people in the working-age population; that is, there’s always a third group of people who are categorized as being “not in the labor

#### 4. Unemployment: Who Wants to Work?



force.” This would be those people who are neither working nor actively seeking work. This category includes stay-at-home parents, discouraged workers (individuals who lost their jobs at some point, looked for work, and gave up actively looking for work), and individuals who cannot work because of their own ill health or caring for someone else.

The employed can be further subdivided into those employed full-time and those employed part-time. Any individual who works between 1 and 34 hours per week is considered to be a part-time worker. About one-sixth of all employed workers—around 25 million individuals—were part-time workers in 2020.

## Calculating the Unemployment Rate

The most commonly used measure of the health of the labor market is the unemployment rate, which is the fraction of the labor force that is unemployed. In 2020, the unemployment rate was 6.7%. In a labor force of 160 million, a 1-percentage-point increase in the unemployment rate would

#### 4. Unemployment: Who Wants to Work?

mean an additional 1.6 million people who would like to work but are unable to find work. At the start of the pandemic, when unemployment rose by nearly 11 percentage points, about 17 million people lost their jobs.

Over time, many of those people returned to work. When people moved back to being employed, the unemployment rate would drop, showing that the labor market was improving. However, other individuals who lost their jobs may have stopped looking for work due to various reasons. They would be categorized as having dropped out of the labor force altogether and would not be factored into the calculation of the unemployment rate. Such omissions can lead to misleading movements in the official unemployment rate.

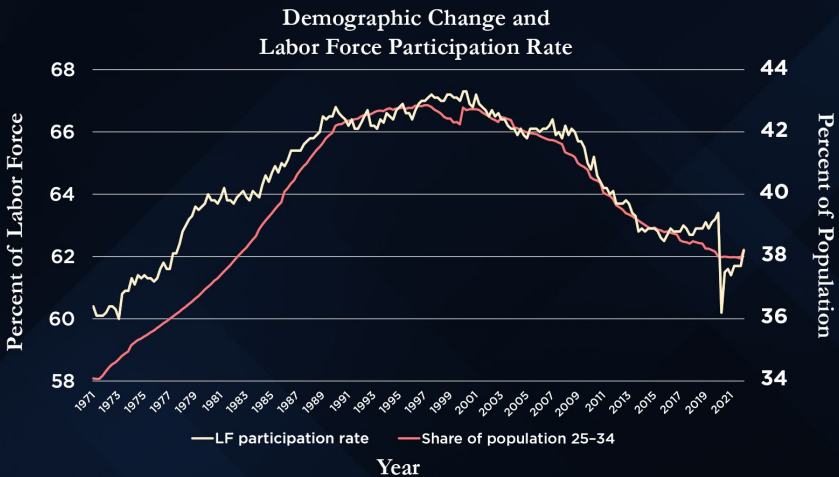
For example, suppose there were 250 million people in the population, and 150 million of those people were in the labor force, of which 7.5 million were unemployed. In that case, the unemployment rate would be 7.5 million over 150 million, or 5%. Now, suppose a large shock resulted in 5 million additional people losing their jobs and becoming unemployed. In that case, the number of unemployed would be 12.5 million. The labor force would remain at 150 million, and the unemployment rate would rise to 12.5 million over 150 million, which is 8.33%. Now, suppose that half of those 5 million newly unemployed workers decided to stop actively looking for work. In that case, the number of unemployed would drop to 10 million. The labor force would also shrink to 147.5 million—the 2.5 million individuals are no longer considered to be unemployed and, hence, are not in the labor force. The unemployment rate would now be 6.8%.

However, the 1.5-percentage-point drop in the unemployment rate in this example did not mean that the labor market improved. During the pandemic, as in many cases, a drop in unemployment could often mean the opposite—people's employment prospects diminished to the point that they no longer actively sought work. Suppose labor market prospects improve and those 2.5 million workers return to look for work; half of them (1.25 million) find jobs, but the other 1.25 million still have to keep looking. This means that the unemployment rate would go up, to 7.5%. Thus, although the labor market has improved, with more people employed, the unemployment rate has worsened. As such, economists also use the labor force participation rate and the employment-to-population ratio to gauge the health of the labor market.

# The Labor Force Participation Rate

The labor force participation rate is the fraction of the working-age population that is in the labor force. It asks what fraction of those who are capable of working are, in fact, working or actively looking for work. For 2020, the labor force participation rate for the United States was 61.5%. When looked at historically, this rate shows long-term trends that overwhelm the short-term fluctuations. There was a steady and sharp trend of increase in the labor force participation rate from the 1960s through the 1990s. Then, a steady decline began in the 2000s, with an acceleration of the decline following the major downturn in 2008/2009.

Part of the change reflects demographic changes: As the baby boom generation came of working age in the 1960s and began entering the labor market, the labor force participation rate rose. Then, as the baby boomers began to hit age 55 in the year 2000, they increasingly began to drop out of the labor force. Currently, the millennials have replaced the baby boomers as the predominant group in the labor force. However, the US Census Bureau projects that the size of the millennial group will be smaller than the size of



#### 4. Unemployment: Who Wants to Work?

the baby boomer group. Thus, the millennials will need to have higher labor force participation rates if the United States is to reverse the rate's declining long-term trend.

Another change underway over that same period reflects changing gender roles and opportunities in the labor market and within households: More women began to enter the labor market beginning in the 1960s, causing their labor force participation rates to rise. Simultaneously, the decline of manufacturing jobs meant that good labor market opportunities in traditionally male-dominated sectors were declining, resulting in reduced labor force participation among men.

A third reason for the decline in labor force participation in recent times is that the negative impact of a temporary recession can have long-lasting impacts on labor force participation. The 2008/2009 recession and the slow subsequent recovery were particularly disruptive to older workers. Many were unable to return to work and were pushed into retirement. Moreover, the COVID-19 recession in 2020 caused especially severe disruptions to women's labor force participation. Many of the jobs that vanished at the start of the pandemic were in service sector jobs that were more likely to be held by women. In addition, women often found themselves bearing a greater burden of home and family care. Whether these women return to the labor force will also have a significant impact on the trend of the labor force participation rate over the next decade.

## The Employment-to-Population Ratio

The employment-to-population ratio is the fraction of the working-age population that is employed. In 2020, this ratio for the United States was 57.4%. Economists typically consider this ratio as the measure that best captures the health of the labor market. It is high when the unemployment rate is low and the labor force participation rate is high.

The employment-to-population ratio tends to fall during recessions and rise during recoveries. It showed a positive trend until 2000 but then fell somewhat in the early 2000s before a precipitous drop in the 2008/2009 recession, declining by about five percentage points. The ratio reveals how damaging the 2008/2009 recession was to the economy. First, it took close

#### 4. Unemployment: Who Wants to Work?

Employment-to-Population Ratio  
in the US Economy 1971–2021



to a decade to recover even half of the decline suffered during the recession. Second, even in the relatively halcyon labor market of late 2019, when unemployment rates were at historically low levels, the employment-to-population ratio still had not recovered to what it had been in the period from the 1990s until the 2008/2009 recession. Finally, the arrival of the COVID-19 recession briefly lowered this ratio to levels not seen in the entire postwar period. The decline was almost 10 percentage points—twice as large as it was in the 2008/2009 recession. This time, the economy made up most of that decline during the first two years of the ensuing recovery.

Going forward, compared to the unemployment rate, the employment-to-population ratio returning to levels experienced before the 2008/2009 recession is a better, more robust indicator of labor market recovery from the substantial negative impacts of the last two recessions. The broader view offered by this ratio has also led to a broader alternative calculation of the unemployment rate: the broad unemployment rate. Recall that if working-age people get discouraged at the prospect of finding a job and stop looking, they are no longer considered to be unemployed. Furthermore, there are some part-time employed

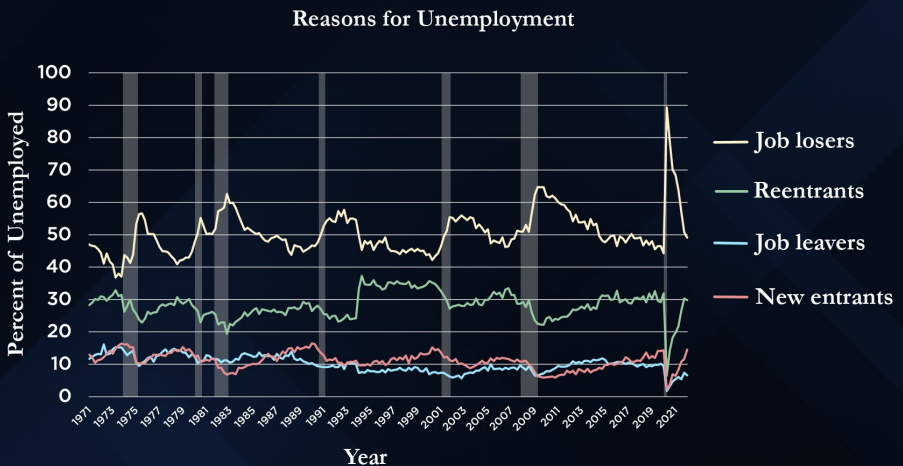
#### 4. Unemployment: Who Wants to Work?

workers who want to have a full-time job but are unable to find one; they are labeled as being “part-time for economic reasons.” The idea behind the broad unemployment rate is to capture how large the sum of all of these groups is relative to the labor force, broadened to include discouraged workers.

By construction, the broad unemployment rate will be higher than the regular unemployment rate, especially in sharp downturns. For 2020, the sharp spike at the start of the pandemic in the broad unemployment rate reached 22.5% compared with the 14.8% reached by the regular unemployment rate. At the end of 2020, the former was 11.7% as compared to the latter of 6.7%.

## Worker Unemployment

There are four broad reasons why people are counted as unemployed. The first is job loss. When a company shuts down or restructures itself, people involuntarily lose their jobs because they are let go by the company. People also lose their jobs because of bad job performance or misconduct. Job loss accounts for about 50% of unemployment. It is particularly pronounced during recessions, most dramatically during the 2020 recession.



#### 4. Unemployment: Who Wants to Work?

The second most common reason for unemployment applies to the group called reentrants. These are workers who left the labor force at some point in the past but are now actively trying to find work again. They typically account for about one-third of all unemployment. However, reentrants often have some choice over when to reenter. Thus, during recessions, potential reentrants are less likely to seek employment and therefore contribute somewhat less to total unemployment.

A third source of unemployment, accounting for roughly 10% to 15%, is “job quits.” This covers workers who leave their jobs voluntarily. Unemployment due to job quits also declines during recessions, as workers tend to hold on to their existing jobs until circumstances become more amenable to a switch in roles. The fourth group of unemployed workers is new entrants to the workforce—accounting for another 10% to 15% of unemployment. Every year, millions of high school and college graduates seek opportunities for work. Recessions can hit new entrants hard, though it also depends on education: Official unemployment can be three times worse for those who don’t finish high school compared to college graduates.

The median length of unemployment has historically been around 8 weeks, indicating that 50% of unemployed people are back at work within a couple of months. However, in the 2008/2009 recession, there was an alarming rise in the length of unemployment, with the median unemployed worker being without a job for as high as 25 weeks. This means that only half of all unemployed people were back at work within six months. Even a decade later, the median point at which half of the unemployed workers found a job was still 10 weeks rather than 8 weeks.

What about all the people who had more difficulty than the median? The mean length of unemployment gives more weight to the outliers, and that average rose to 40 weeks during the 2008/2009 recession. Even a decade later, the mean was still averaging 20 weeks of unemployment. Economists pay particularly close attention to those whose unemployment spells last more than 26 weeks. Such long spells are likely to be associated with a deterioration of skills and connections that will make a successful return to employment less likely.

## READING

Aaronson, Stephanie. “What Does the Unemployment Rate Measure?” Brookings, February 18, 2021. <https://www.brookings.edu/blog/up-front/2021/02/18/what-does-the-unemployment-rate-measure/>.

American Association of University Women. “The Simple Truth about the Gender Pay Gap.” Fall 2018. <https://files.eric.ed.gov/fulltext/ED596219.pdf>.

Deryugina, Tatyana, Olga Shurchkov, and Jenna Stearns. “Covid-19 Disruptions Disproportionately Affect Female Academics.” *AEA Papers and Proceedings* 111 (May 2021): 164–168.

Goldin, Claudia. *Career and Family: Women’s Century-Long Journey toward Equity*. Princeton University Press, 2021.

Krueger, Alan B. “Where Have All the Workers Gone? An Inquiry into the Decline of the US Labor Force Participation Rate.” *Brookings Papers on Economic Activity* no. 2 (2017): 1.

Whitmore, Diane Schanzenbach, and N. Ryan. “The 51%: Driving Growth through Women’s Economic Participation.” Brookings, October 19, 2017.

# 5

## HOW IS MONEY CREATED— AND WHY?

**T**his lecture is all about money: why and how it was created, how it has continued to evolve, and what roles it plays in the economy, especially in terms of the Federal Reserve. Money is essential for keeping a robust flow of economic activity going. Finding easier ways for people to use this medium of exchange—debit cards, credit cards, Venmo, Apple Pay—has become a key feature of the modern economy. Economists think of money as serving as a medium of exchange, a store of value, and a unit of account in the economy.



## What Is Money?

In ancient Mesopotamia, the term *shekel*, which meant one-third of an ounce of silver, came into use with the first known form of money. Gold, silver, and bronze coins were used in ancient civilizations in China, India, Egypt, and Asia Minor. These coins helped increase economic activity because they could be carried around easily, were durable, and could be easily divided into smaller units. Taxes could also more efficiently and easily be collected from the populace in the form of coins, which could be used to attract more soldiers for armies to conquer new lands or defend existing ones.

Furs in cold climates and salt along inland trade routes are examples of commodities that were also used as money because of their intrinsic value. In Eastern Europe, behind the Iron Curtain, vodka became a valued commodity currency since it could be exchanged for Western goods. In prisons today, packets of mackerel filets are used to trade goods and services.

The medium of exchange refers to the willingness of someone to accept money in exchange for the good or service they have to offer. The store-of-value aspect is also important. A farmer can sell her produce for money after the harvest and use the money to buy things for her family during the growing season. Storing the harvest is much harder because of the cost and natural spoilage. One reason why foil packs of mackerel are valued in prison is because they have a longer shelf life than other goods that one can buy at a commissary. The store-of-value aspect is closely linked to the medium-of-exchange aspect. Economists see the store-of-value function illustrated in the currencies of countries where inflation rates are extremely high. People try to get rid of the paper currency as soon as they can because the loss of value makes it undesirable as a medium of exchange.

The unit-of-account feature of money is that the money can be divided into smaller amounts. However, not all commodities used as money can easily be divided this way. For example, the practice of using livestock as money has endured despite it not being a good unit of account.

Beginning around the 1800s, paper notes began to be used as money and soon began to overtake the use of coins. Paper money was easier to produce, transport, and divide but had the obvious drawback of not being a

## 5. How Is Money Created—and Why?



commodity—there was no intrinsic value of the note itself. For people to be comfortable accepting paper money as a medium of exchange, they needed to be convinced that it retained its store-of-value feature. To this end, the amount of paper currency that governments could issue was linked, by law, to the

supply of a commodity, such as gold

or silver. In the 19th and early 20th centuries, many countries were on a gold standard, which meant that the price of gold in terms of paper money was fixed by the government. In the United States, for example, the price of gold was set to \$21 per ounce. This meant that the government agreed to buy and sell gold at that price in exchange for paper currency. However, the gold standard collapsed in 1933.

## Current Currency

The Bretton Woods system that began in 1945 was based on gold and the US dollar. At the time, the United States controlled two-thirds of the world's gold. Technically, the US dollar was fixed to gold at a price of \$35 per ounce. However, this rate became increasingly difficult to maintain over the 1960s as the money supply of US dollars increased. This created a mismatch between the price at which gold could be directly bought or sold on the world market and the rate at which dollars could be converted to gold through the Bretton Woods system. The United States ended dollar convertibility to gold in 1971. Other countries severed all links between their money and the dollar under Bretton Woods by 1973. As a result, the currency used today—coins and paper money—is fiat money as opposed to commodity money: It has value simply because people value it as a medium of exchange.

## 5. How Is Money Created—and Why?

In the modern economy, the bulk of what people consider to be money comes in the form of deposits at financial institutions. People use various instruments, such as debit cards or Venmo, to access this money directly for transactions. Money gets categorized by the Federal Reserve according to how liquid it is—that is, how easy it is to use in transactions. Prior to May 2020, the official categorizations were M1, which comprised currency and checking account balances held in banks by individuals and firms as well as traveler's checks; and M2, which comprised everything in M1 plus savings deposits, time deposits (e.g., certificates of deposit, which cannot be accessed for a specific period of time), and other deposits where check writing is limited. Items in M2 that are not in M1 are not as liquid as those assets in M1.

### Categorized as M1

---

Currency  
Checking account balances  
Traveler's checks

### Categorized as M2

---

M1  
Savings account balances  
Certificates of deposit  
Money market mutual funds

Financial innovation continually blurs the line between what is M2 and what is M1. For example, in mid-2020, the Federal Reserve removed the restriction of a maximum of six withdrawals a month that it had imposed on savings accounts. As a result, people could effectively access the money in certain types of savings accounts just as easily as they could access their checking account balances or the currency in their wallets. Thus, savings account balances that used to be considered as M2 were now labeled as M1. Prior to May 2020, M1 was around \$4 trillion, and M2 was around \$16 trillion. After the recategorization, M1 and M2 were almost identical at around \$20 trillion. Given that nominal GDP that same year was around \$22 trillion, effectively, about 20 trillion units of money were supporting \$22 trillion worth of economic activity. Currency is only about 10% of the value of M2; around \$2.5 trillion worth of notes and coins circulate in the economy.

# Cryptocurrency

Invented in 2009, Bitcoin is a decentralized digital currency, or cryptocurrency. The allure of cryptocurrencies is twofold: First, they are a digital asset that can be transferred electronically and securely using cryptographic methods. Second, cryptocurrencies are not created by a central bank or government. This decentralized nature can make them an attractive alternative in countries whose central banks have a bad reputation for printing too much currency, resulting in high inflation that wipes out the purchasing power of money. Critics of cryptocurrencies often point to a third aspect for their popularity: They can be transferred outside of the financial system and away from legal and tax scrutiny.

Cryptocurrencies have fluctuated greatly in value. Bitcoin went from about \$600 in December 2013 to about \$19,800 in December 2017 and back to \$3,300 in December 2018. It climbed into the \$60,000 range in 2021 and fell back into the \$20,000 range in 2022. At peak market value, all Bitcoins combined were worth more than \$1 trillion—roughly the M1 of Canada or Australia. These spectacular increases, paradoxically, make cryptocurrencies less likely to serve as money rather than as a speculative asset. Their allure of being too good as a store of value can lead to them becoming less used as a medium of exchange.

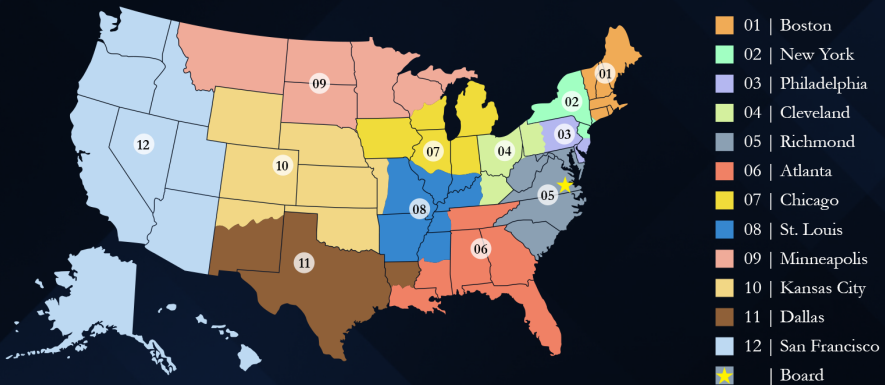
For example, a cryptocurrency called the Shiba Inu was created in 2020 based on an internet meme. Its value boomed in 2021, increasing 330,000-fold between January and December. Thus, \$100 worth of Shiba Inu coins purchased in January of 2021 would have been worth \$33 million in December of 2021. Shiba Inu coins, however, were not used to make everyday purchases. The holders of these coins kept them to cash them in later. However, not all cryptocurrencies are as volatile as the Shiba Inu. So-called stablecoins, such as Tether, are tied in value to an existing paper currency and (in theory) backed by reserves of traditional currency. As such, volatility alone will not determine whether the use of cryptocurrencies increases over time.

Regardless of whether cryptocurrencies will ever supplant paper currency, it is clear that central banks have taken a keen interest in the rise of electronic currency. Central banks, which control the supply of fiat currency, are considering—and, in some cases, issuing—their own electronic money to be used by individuals in the economy.

## The Federal Reserve

The Federal Reserve—or the Fed—has served as the central bank of the United States since it was established in 1913. It has two arms. The Board of Governors, located in Washington DC, consists of seven people appointed to nonrenewable 14-year terms by the US president and then confirmed by the Senate. One of the governors is appointed by the president as chairman of the board. Moreover, there are 12 regional Fed Banks distributed around the country. These banks monitor local economic conditions and act as a “bank to banks” in their region. Effectively, they process checks, facilitate electronic payments between banks, and act as overseers of the banks in the district.

The 12 Federal Reserve Districts  
Federal Reserve Banks



The key decision maker with regard to the money supply is the Federal Open Market Committee (FOMC), which consists of the Board of Governors, the president of the New York Fed, and four rotating presidents of the Fed regional banks. The FOMC meets about eight times per year and decides whether to change the supply of money, and by how much, to achieve its monetary policy goals for the US economy. These goals were laid out by Congress “to promote maximum employment and stable prices in the US economy.”

## 5. How Is Money Created—and Why?

The actions of the Fed to increase the supply of money are transmitted to the economy via banks. The Fed plays a critical role in the banking sector to protect the stability of the financial system and the soundness of individual financial institutions. To this end, it uses several tools. Banks are typically required to keep a certain fraction of their customers' deposits, called the required reserve ratio, at the Fed. Banks are also allowed to borrow money from the Fed if they need money for some reason, such as an unexpected increase in customer withdrawals. The rate of interest at which the Fed lends money to banks is the discount rate. Banks act as intermediaries between savers and depositors (people who want to earn interest on the money they do not intend to use currently for consumption) and investors and borrowers (people who are willing to pay interest to borrow money to undertake spending or invest in projects). Banks make money by charging investors and borrowers a higher interest rate than they pay to the savers and depositors.

The bank's liabilities are its deposits (money that belongs to its customers) and its debt (money that belongs to its creditors). Its assets are the loans it has made to clients or to other banks, any financial securities such as stocks and bonds it holds, and its reserves. Reserves are deposits that the bank chooses to hold on to rather than lending out or investing in securities. They can be held either in the bank's own vault (vault cash) or at the central bank. These reserves can take the form of required reserves, which are the fraction of deposits that the Fed mandates that banks hold. However, banks will sometimes hold more than what is required of them at the Fed, or excess reserves. For example, say a bank holds \$1 billion in deposits and owes \$250 million to its creditors. Its total liabilities are \$1.25 billion. It has \$700 million worth of loans that it has made to others and owns \$550 million worth of stocks and bonds. It has decided to hold no reserves in excess of the required reserves, which are 10% of its deposits, meaning \$100 million.

The difference between assets and liabilities is called bank capital. Essentially, if the bank were to be folded up today, bank capital represents what would be left over for the shareholders and owners of the bank after all its assets were sold off and all its liabilities paid off. Bank capital represents the initial capital that owners put into the bank plus any operating profits that the bank has retained over the years. When you read about a bank failing, this means

## 5. How Is Money Created—and Why?

that its capital has fallen to below zero—its liabilities exceed its assets. When that happens, regulators will shut down the bank and transfer its assets and liabilities to another bank while making sure that depositors are made whole.

In normal times, the Fed changes the money supply in one of three ways. The primary way is via open-market operations: buying or selling government bonds. If the Fed wants to increase the money supply, it will inject money into the financial system to buy government bonds. If it wants to decrease the money supply, it will sell government securities from its existing collection in exchange for cash and pull that money out of the financial system. US government bonds represent a safe asset that the Fed can buy, hold, and sell without worrying about substantial price fluctuations in the underlying asset value.

A second way to change the money supply is by changing the discount rate that the Fed charges banks in exchange for lending them money. This way, the Fed can make it easier for banks to borrow money. This increased borrowing leads to more money circulating in the economy. Conversely, a higher discount rate will make it more costly for banks to borrow from the Fed. This decreased borrowing leads to less money circulating in the economy.

Finally, if the Fed wants to raise the supply of money, it can lower the required reserve ratio. This will permit banks to keep fewer reserves at the Fed and allow them to make more loans. Conversely, if the Fed wants to lower the supply of money, it can raise the required reserve ratio. This forces banks to keep more reserves at the Fed and reduces the amount of loans they can make.

Consider an example where the Fed decides to inject \$100 million into the economy via an open-market operation. Suppose that the Fed buys government bonds directly from a bank. In effect, it puts \$100 million into the bank's reserve account at the Fed in exchange for the bonds. The bank's holdings of financial securities fall by \$100 million, from \$550 million to \$450 million. Its excess reserves rise from \$0 to \$100 million. The bank now has to decide how to use these additional reserves. Notice that the open-market operation did not change the bank's financial health (i.e., its capital did not change), nor did it impact how much the bank was required to hold in reserves (10% of deposits). Thus, the bank is free to lend out its additional reserves rather than keeping them in excess reserves earning little to no interest. When banks lend out the reserves, the supply of money in the economy increases.

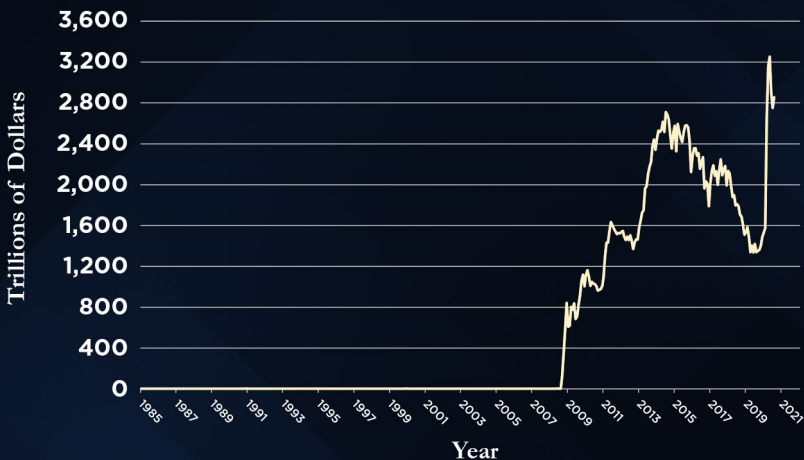
## 5. How Is Money Created—and Why?

However, when there is turmoil in the financial markets and/or the banking sector, banks may become more concerned about making loans to customers and may hold on to the excess reserves instead of lending all the new deposits out. This means that the Fed would need a larger open-market operation to achieve its goals. Yet the way the Fed has operated since 2008/2009 does not fit neatly into the story just described.

# Ample Reserves

At the end of December 2021, banks held almost \$4 trillion in reserves at the Fed, even though the Fed has no longer required banks to hold reserves since March 2020. This sharp increase in excess reserves began after the global financial crisis in 2008/2009. Jane Ihrig and Scott Wolla, two economists at the Fed, characterize this period as one of “ample reserves.” In this world, the Fed has less ability to change the money supply in the economy using traditional open-market operations. When banks have so much in excess reserves, the Fed can’t impact their willingness to lend by adding liquidity into the system.

Excess Reserves Held by Banks in the United States



## 5. How Is Money Created—and Why?

How did this world come about? Consider what happened in 2008/2009 when the housing market imploded and many banks fell into financial difficulty because loans and assets related to the housing market plummeted in value. When an asset falls in value, holding liabilities fixed, the bank's capital falls. If the capital shrinks further, the bank becomes more likely to hold on to the excess reserves rather than risk making more loans into the housing markets, for example.

Furthermore, the Fed helped banks reduce some of their uncertainty by offering to buy up financial securities that banks were reluctant to hold. In such a world, banks quickly become awash in reserves. The Fed cannot influence the amount of money circulating in the economy simply by using open-market operations because reserves are not a scarce resource. It has to adapt to this world of “ample reserves.”

The Fed created unprecedented amounts of M1 in the economy as it dealt with the economic troubles emanating from the global financial crisis of 2008. These problems lingered for a decade and were soon replaced by a new global economic shock: the COVID-19 pandemic. During this 15-year period, the Federal Reserve purchased more than \$7 trillion of assets; that is, more than \$7 trillion worth of reserves have been introduced into the banking sector.

### READING

Goldstein, Jacob. *Money: The True Story of a Made-Up Thing*. Atlantic Books, 2020.

Lowenstein, Roger. *America's Bank: The Epic Battle to Create the Federal Reserve*. Penguin Random House, 2016.

Popper, Nathaniel. *Digital Gold: Bitcoin and the Inside Story of the Misfits and Millionaires Trying to Reinvent Money*. Harper Collins, 2016.

The Federal Reserve. “Currency & Coin.” Accessed December 12, 2022. <https://www.federalreserve.gov/faqs/currency-coin.htm>.

———. “Structure of the Federal Reserve System.” Accessed December 12, 2022. <https://www.federalreserve.gov/aboutthefed/structure-federal-reserve-system.htm>.

# 6

## IS THERE SUCH A THING AS TOO MUCH MONEY?

**G**iven how essential money is to economic activity, a central bank such as the Fed can surely help increase economic activity by printing more money. Thus, why not expand the money supply dramatically? The classic answer is that too much money chasing too few goods leads to inflation. This lecture will dive deeper into the question of how much is too much when it comes to the appropriate amount of money that an economy needs to enhance stability and economic activity.



## The Quantity Equation

Economists use the quantity equation to better understand the link between the quantity of money and the rise in prices. Since money is used for transactions, the equation looks at the amount of money ( $M$ ) multiplied by the velocity of money ( $V$ ), which is the number of times that each unit of money is used for making purchases of new goods and services in a given period. The quantity of money times the velocity should be equivalent to the value of production in the economy, which is nominal GDP. Note that the velocity is calculated from GDP and the money supply using this relationship. For instance, in June 2021, the supply of M1 in the US economy was \$19.2 trillion, and nominal GDP was \$22.74 trillion. You can use  $MV = PY$  (where  $P$  is price level and  $Y$  is real GDP) to calculate  $V = 1.19$ . This means that on average, a dollar of M1 was used roughly 1.19 times for economic transactions during the second quarter of 2021.

$$\begin{array}{c}
 \text{Quantity of Money} \\
 M
 \end{array}
 \times
 \begin{array}{c}
 \text{Velocity of Money} \\
 V
 \end{array}
 =
 \begin{array}{c}
 \text{Price Level} \\
 P
 \end{array}
 \times
 \begin{array}{c}
 \text{Real GDP} \\
 Y
 \end{array}$$
  

$$\begin{array}{c}
 \text{Quantity of Money} \\
 M
 \end{array}
 \times
 \begin{array}{c}
 \text{Velocity of Money} \\
 V
 \end{array}
 =
 \begin{array}{c}
 \text{Nominal GDP} \\
 P \times Y
 \end{array}$$

The growth of the money supply plus velocity growth (how fast the use of money is growing) should equal the growth rate of prices, which is inflation, plus growth in real GDP. This is the key relationship economists use to understand the link between money growth and inflation. The left side of the equation is a measure of two ways the amount of money used for transactions can grow: Either the stock of money is increasing or each unit of money is being used more often for transactions. The right side is a measure of two ways the value of production can grow: The quantity of goods being produced can be rising, or the price of those goods can be rising. Since the money being used for transactions is purchasing the produced goods, the two sides have to be equal by definition.

Thus far, this is an accounting identity rather than an economic theory about how money growth is related to inflation. Economists do not independently observe velocity in the economy. Velocity is indirectly calculated as the value that makes  $M \times V$  equal to  $P \times Y$ , so this equation alone does not give any particular insight into how money growth may affect inflation in the economy.

# The Quantity Theory of Money

The above equation can be transformed into the quantity theory of money via a set of assumptions about the equation's various pieces to derive predictions about the relationship between money growth and inflation. The monetarist school of economics used this theory to argue that inflation was primarily the result of central bankers trying to produce too much of what the bankers perceived to be a good thing for the economy. The first assumption here is that real GDP growth is not driven by changes in the growth rate of the money supply.

If people could change the growth path of an economy by printing more money, then economic growth would hardly be a challenge. In economics, this idea is called monetary neutrality, and it is imposed on the quantity equation. The assumption does not preclude the possibility that over short periods of time—when economies are in the middle of a financial crisis, for example—money growth can facilitate more spending and be positively linked to real GDP growth.

Next, consider possible assumptions about velocity and see what predictions follow from each assumption. Begin by assuming that the velocity of money is a constant. Although unrealistic, this would mean that the growth rate of velocity would be zero. In this case, the quantity equation would simplify to say the following: money growth + 0 = inflation + real GDP growth. Now, suppose the long-term growth rate of real GDP is 3%. This is assumed not to be affected in the long term by money growth. Under these circumstances, you have the following equation: money growth = inflation + 3%, or inflation = money growth - 3%.

Assume a central banker sets money growth to be 3%. She maintains that 3% money growth for another year. However, she then decides to increase money growth to 8% for the next year. After that experiment, she returns to 3% money growth rate for a year. Finally, she experiments with 1% money growth

## 6. Is There Such a Thing as Too Much Money?

for the fifth year. Given your assumptions that velocity is not changing and real GDP is growing at 3%, the quantity equation implies that inflation is equal to money growth minus 3%. You obtain striking results: First, there is no inflation (prices are stable) when money growth equals the real GDP growth rate of 3%. Second, there is inflation when money growth exceeds the real GDP growth rate. Third, there is deflation when money growth is less than the real GDP growth rate. Fourth, the rate of inflation will change one-for-one with the rate of money growth: An  $x$ -percentage-point change in money growth will change inflation by  $x$  percentage points.

In a model where velocity is assumed to be constant, inflation occurs when there is proportionally more money chasing goods. Deflation occurs when there is proportionally less money chasing goods. According to this simplified assumption, a central banker who commits to a steady (low) growth rate of money supply will deliver a steady (low) rate of inflation.

However, velocity is not constant. How would the outcomes for the central banker change if velocity grew at a constant rate of 2%? Now, the relationship between money growth, velocity, inflation, and GDP is as follows: money growth + 2% = inflation + 3% (where you are again assuming real GDP growth is constant at 3%). If you rearrange this, inflation will be 1 percentage point higher than money growth: inflation = money growth - 1%.

Consider how the results in this case compare to those in the previous case: First, for both cases, the rate of inflation is higher than it was under the assumption of constant velocity. Given that velocity was growing at 2%, inflation would be 2 percentage points higher. Second, there is inflation even when money growth equals real GDP growth. Third, the rate of inflation will change one-for-one with the rate of money growth. Thus, under a constant growth rate of velocity, inflation occurs when the growth rate of money used for purchases of goods and services exceeds the growth rate of production.

The growth rate of the amount of money being used for transactions depends on the growth rate of the money supply and the velocity growth, both of which reinforce each other. Thus, if you have velocity changing at a constant rate, you can say that a central banker who wants to keep inflation low and stable can achieve this goal by keeping money growth low and stable. She would simply need to keep money growth lower than in the case of unchanging velocity.

## Changes in the Velocity of Money

Change in the velocity of money can occur for various reasons. For example, velocity will rise as interest rates rise and fall as interest rates fall. Suppose a one-year certificate of deposit (CD) is offering an 8% interest rate. A one-year CD is not considered to be M1 because you can't easily access that money for a year. If it is offering you an 8% return, you will think hard about what you keep in checking balances or cash because the opportunity cost of keeping excessive balances in liquid accounts is high. Since what you keep in M1 is more likely to be used in the purchase of a good or a service, velocity will rise.

Conversely, when interest rates are low and that one-year CD is paying you 0.25%, you are less concerned about the opportunity cost of keeping your money in checking accounts or as cash. These have the added benefit of being useful for economic transactions. If you hold more money than what you will likely need for transactions over the year, then velocity will fall since each unit of money is less likely to be used in a transaction. Thus, you need to relax the assumption that velocity grows at a fixed rate. Then, the results become even more complicated. Now, you are saying that velocity will grow at some rate but that that rate can change from period to period.

Inflation can change even when money growth rates are stable. Rates of inflation can be higher or lower than under the case where velocity growth is constant. There is inflation even when money growth equals real GDP growth. The rate of inflation will not change one-for-one with the rate of money growth. An  $x$ -percentage-point change in money growth can change inflation in unpredictable ways, depending on how velocity growth is changing. However, note that if the growth rate of the money used for transactions (the sum of velocity growth and money growth) exceeds output growth, then there will be inflation. If the growth rate of the money used for transactions is less than output growth, there will be deflation.

If the Fed wants to keep inflation at zero in the face of fluctuating velocity growth, a constant money growth rate will not suffice—inflation will move around quite a bit. With a changing velocity growth rate, a good central banker will have to constantly adjust money growth policies in the face of changing economic circumstances to keep inflation stable. For example, if people are spending less money after a recession, then the Fed needs to

## 6. Is There Such a Thing as Too Much Money?

provide more money growth into the economy to avoid deflation. If people are spending money at a faster rate due to a receding pandemic, for example, then the Fed needs to reduce money growth to avoid inflation.

The key insights here are that a good central banker cannot simply be an automaton increasing money supply at some constant rate and that understanding how to adjust money growth in the face of changes in velocity is important. In fact, the growth rate of money itself can cause the growth rate of velocity to change. A rapid growth rate of money can cause inflation to spiral out of control. This is how hyperinflations can sometimes occur in an economy such as Zimbabwe's.

# Hyperinflation

Suppose there is an economy where inflation is stable at 2%, real GDP growth is stable at 3%, and velocity is growing at a rate of 1% per year. Money growth plus velocity growth of 1% equals inflation of 2% plus real GDP growth of 3%. You can then use the quantity equation to determine that the central banker has been increasing the money growth by 4% a year in this economy.

Suppose that the central bank increases money growth by 5 additional percentage points, for a new money growth rate of 9%. Then, if you assume velocity growth stays constant, inflation becomes 5 percentage points higher and will equal 7%. As inflation rises, money begins to change hands much more quickly. Thus, you would expect an increase in the growth rate of velocity—to 5%. The combination of increased money growth and velocity growth implies an even more rapid increase in inflation, which would now rise to 11%. As prices rise by double-digit levels, unhappiness increases. The only way to keep the people placated is to pay them more so that they can afford to buy expensive goods. However, this requires printing more money. Thus, you increase money growth by another 10 percentage points. This increase in the rate of money growth again leads to a further rise in inflation, to 21%.

Inflation has now increased 10-fold, from the original 2% to a new value of 21%. As the purchasing power of money diminishes rapidly, people start spending money as soon as they get their hands on it. This increases velocity growth even more and fuels more inflation. Soon, the process escalates, and 100% inflation becomes common. In a few notable cases, such as Zimbabwe in the 2000s, inflation in the billions of percent becomes a distinct possibility.

## 6. Is There Such a Thing as Too Much Money?

The quantity theory can explain why not all large-scale changes in the money supply lead to increases in inflation. Consider the behavior of M1 in the United States after the financial crisis and recession of 2008/2009. Money growth from 2009 to 2015 was consistently high, increasing between 5% and 25% a year. However, inflation stayed stable. Prices rose by less than 2% a year, undershooting the Fed's target rate of inflation of 2%.

The quantity theory helps you reconcile these two pictures: The reason that rapid increases in money growth did not fuel inflation is because the growth rate of velocity must have fallen. The usual calculations can show that velocity growth was strikingly negative during those years. In fact, the Fed increased the money supply so much because it knew that the financial crisis and recession were causing sharp slowdowns in spending patterns. If velocity growth was falling, the Fed would need to raise money growth rates to keep inflation stable.

## Hyperdeflation

Recall the emergence of cryptocurrencies as a potential alternative to traditional currencies. The runup in values of cryptocurrencies, as seen in 2020 and 2021, may paradoxically lead to them being less used as money in the economy. This is hyperdeflation—the amount of cryptocurrency needed to buy a particular good or a service decreases dramatically.

Apply the quantity equation to the cryptocurrency context. Money growth will be the increase in the supply of cryptocurrency; inflation will be the increase in prices measured in cryptocurrency terms. Real GDP growth will be the increase in the types of goods and services that can be bought with cryptocurrencies. By construction, the rates of money growth for some cryptocurrencies, such as Bitcoin, are designed to slow down over time. Furthermore, as cryptocurrencies become more prominent, more and more companies may be willing to accept cryptocurrencies. Thus, the last term on the right keeps increasing.

Assume that velocity growth is constant. Since the rate at which people are spending cryptocurrency falls short of the goods and services that are being offered in exchange, the inflation rate must be falling. The cryptocurrency is becoming more valuable over time. This leads to more people being willing

## 6. Is There Such a Thing as Too Much Money?

to exchange goods and services for it. However, those holding cryptocurrency become much less willing to exchange it for goods and services. Thus, cryptocurrency velocity growth falls. Soon, a hyperdeflationary cycle will set in. As the deflation rate rises, with fixed supply and rising numbers of goods and services offered in exchange for cryptocurrencies, people will hoard cryptocurrency, causing further deflation, and so on. This won't be true for those cryptocurrencies that can increase their supply. However, if the supply of a cryptocurrency can be increased without bound, what makes it superior to an electronic currency issued by a central bank?

There's an idea, from David Hume and refined by Milton Friedman, that inflation comes from "too much money chasing too few goods." That's a good start, but it's also incomplete. Using the quantity theory of money, you saw that the assumption of a tight link between money growth and inflation growth depends on a constant rate of velocity growth. In reality, velocity growth can vary. There can be various outcomes for inflation resulting from a given path of money growth. That's why inflation stayed low despite the Fed increasing money growth by double-digit rates following the 2008/2009 recession. The Fed understood that falling velocity growth required a higher rate of money growth to ward off deflation. However, in Zimbabwe, the central bank increased the rate of money growth in a setting where velocity growth was also increasing. This combination of money growth and velocity growth led to higher inflation. This in turn led to further increases in velocity growth rates and spiraling inflation.

By contrast, a decentralized cryptocurrency has traditionally had built-in limitations on money growth. Thus, any decline in its velocity poses the danger of deflation. A good medium of exchange depends on not losing too much value but also not gaining too much value. Being a good manager of the medium of exchange requires paying careful attention to velocity rather than merely keeping pace with the production of goods and services. If people are reticent about spending money, the government may have to print more money to keep inflation stable and avoid deflation. If people are spending money quickly, the government must slow down the rate at which it prints money to keep inflation stable.

## READING

- Federal Reserve Bank of Dallas. “Hyperinflation in Zimbabwe.” *Globalization and Monetary Policy Institute 2011 Annual Report* (2011): 2–12. <https://fraser.stlouisfed.org/title/annual-report-globalization-monetary-policy-institute-6346/2011-annual-report-606582>.
- Friedman, Milton. “Money and Inflation: Milton Friedman Speaks” (video lecture). Hoover Institution, November 7, 1977. <https://miltonfriedman.hoover.org/objects/57219/money-and-inflation>.
- . *Money Mischief: Episodes in Monetary History*. HMH, 1994.
- Samuelson, Robert J. *The Great Inflation and Its Aftermath: The Past and Future of American Affluence*. Random House Trade Paperbacks, 2010.
- Sargent, Thomas J. “The Ends of Four Big Inflations.” *Inflation: Causes and Effects* 1 (1982): 41–98.

# 7

## TWO INTEREST RATES TO RULE THEM ALL?

**T**here are millions of interest rates in an economy. There are savings rates and bond rates and loan rates on everything from mortgages, student loans, and car loans to credit cards and payday loans. Each of these can impact the decisions of consumers, firms, and the government to spend, save, borrow, and invest. In this lecture, you will learn how these different rates can be understood with the help of the federal funds rate and the yield on US Treasury bonds.



## Understanding Interest Rates

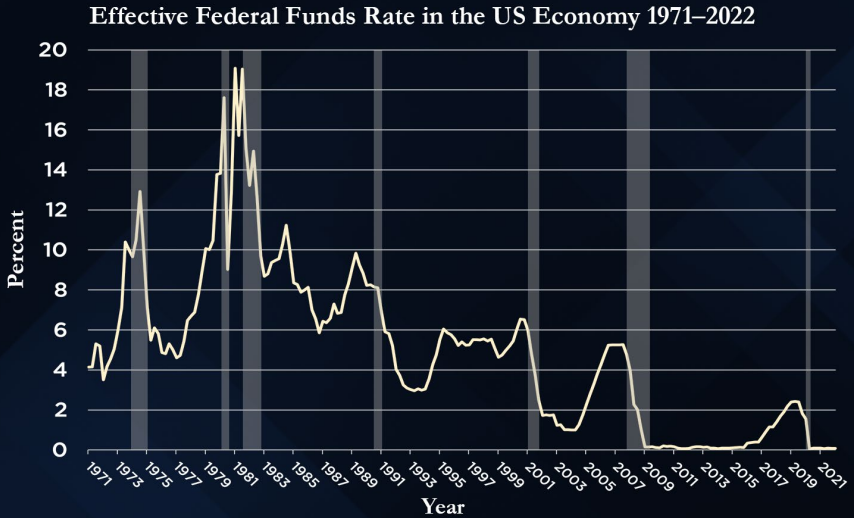
Think of an interest rate as being simultaneously the price of borrowing money and the gain from lending money. Interest rates are typically expressed on a percentage-per-year basis. For example, if the interest rate is 5%, someone who borrows \$100 today will have to pay back the original \$100 plus 5% of that \$100, for a total of \$105, in a year's time. This means that a dollar in the future is worth less than a dollar today. How much less a future dollar is worth depends on the interest rate.

Suppose you can earn a guaranteed 5% interest rate over the next five years by putting your money in the bank. If you put \$1,000 into the bank, you will have \$1,050 next year. If you leave that \$1,050 in the bank for another year, you will earn a further \$52.50, taking the balance to \$1,102.50. After five years, you will have \$1,276.28. Thus, the future value of \$1,000, given an interest rate of 5%, is \$1,276.28.

Higher interest rates increase the future value of money relative to the present. This encourages people to save more and borrow less now. Conversely, lower interest rates increase the present value of money, which encourages people to save less and borrow more now.

What determines the interest rate differential you are charged when you take out a 15-year loan to buy a new fishing boat compared to the rate you will be charged for a 15-year mortgage loan? The boat loan will invariably come with a higher interest rate. Why? The home loan is less risky to the lender because it offers better collateral. If you default on your home loan, they can repossess your land and your house. However, if you default on your boat loan, they inherit a used boat, which is much harder to sell and may have depreciated substantially in value. The less risk there is in a loan, the lower the interest rate.

What about differences in rates for identical loans taken out by different borrowers? The interest rates at which you and your neighbor can each take out a 30-year mortgage to buy a house depend on your respective credit ratings. The mortgage company will check credit scores and charge higher interest rates to individuals who have lower credit scores. The higher the risk, the higher the interest rate.



## The Federal Funds Rate

A few key interest rates have such low risk that they are regarded as risk-free interest rates. The most important of these is the federal funds rate. This is the interest rate that banks charge one another for borrowing money overnight on the federal funds market. Since the 1990s, the target for this rate has typically been 2% to 5%, but it was lowered to 0% in response to the financial crisis of 2008. It was lowered to 0% again at the start of the pandemic in 2020.

The federal funds rate underlies most of the higher rates that ordinary borrowers care about, including the prime rate, which is the rate that banks charge their largest and most creditworthy customers. In recent years, the gap between the federal funds rate and the prime rate has tended to be about 3%.

Suppose a top company needs a one-year loan to buy additional inventory for its production line. The company's CFO asks the bank for a loan. If the federal funds rate is 2%, then the banker will charge the company an interest rate based on the higher prime rate of 5% for the loan. However, if the federal

funds rate is 1% instead of 2%, the banker may be willing to make that same loan for 4%. Now, suppose a small-business owner approaches the bank for a loan. If the federal funds rate is 2% and the bank's best client is being charged 5%, this new client may be charged 6.75%, for the loan. The additional 1.75 percentage points are called a risk premium.

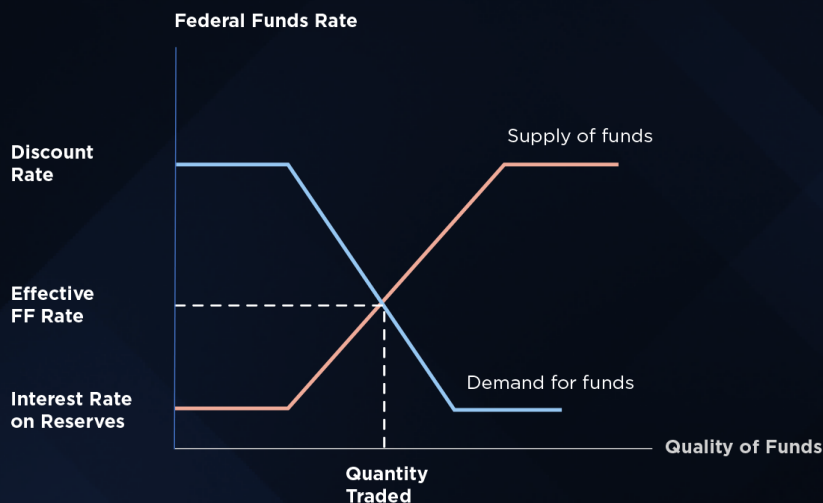
Consider how other government interest rates typically move with the federal funds rate. Here, the other interest rates are the interest rates that the US government pays when it borrows money for three months, six months, or one year. These rates are closely related. The Fed does not directly set the federal funds rate. It is determined by supply and demand conditions in the federal funds market for overnight loans between banks. The Fed can and does intervene to influence the rate at which banks lend money to one another by setting a target rate for the federal funds rate. It then moves to add or remove funds from the federal funds market to ensure that its target rate is achieved.

## Determining the Federal Funds Rate

The demand for funds in the federal funds market is downward sloping because the higher the cost of borrowing, the less banks want to borrow. However, the rate has a ceiling: the discount rate. This is the interest rate that the Fed charges to banks for lending them money. Thus, the interest rate in the federal funds market will not rise above that ceiling. Banks can borrow from the Fed at the discount rate rather than borrowing from each other at a higher rate. The federal funds rate also has a floor. This is the interest rate the Fed pays on reserves (the excess reserves) that banks hold at the Fed. Thus, the federal funds rate will never fall below that floor because no bank will lend at that rate. Banks can earn a higher interest rate by keeping their money at the Fed.

When the Fed wants to make a policy change, it announces that it is changing its target rate for the federal funds market. It then uses open-market operations—buying or selling US Treasury securities from banks—to increase or decrease the amount of funds that banks can trade in the federal funds market. The Fed takes one of three policy stances: ease, tighten, or no change. When it wants economic activity to pick up, perhaps because there is a recession looming, it wants to ease its policy stance. Thus, it lowers its target rate for the federal funds market. It will then conduct an open-market operation to purchase securities from banks. Fed purchases will leave the

## 7. Two Interest Rates to Rule Them All?



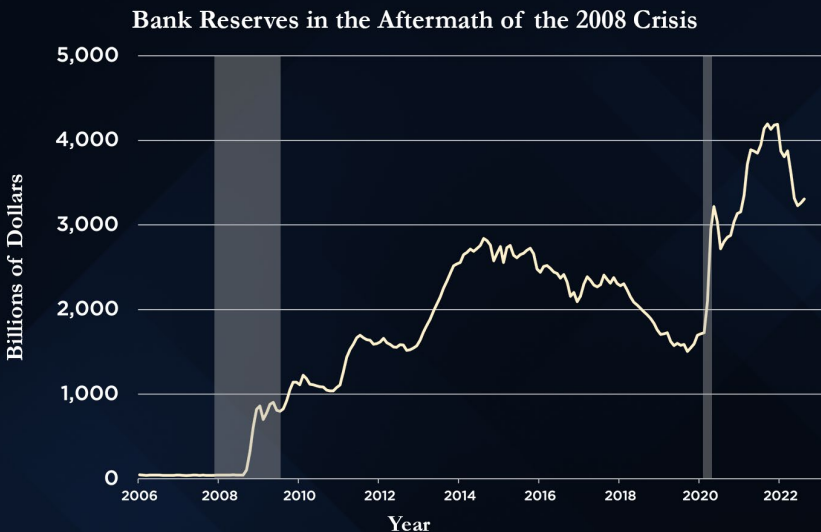
banks with more reserves to lend on the federal funds market, meaning the supply curve will shift out. Thus, the supply of funds in the federal funds market will increase, and the federal funds rate will fall. The Fed will calibrate the size of the open-market operation to deliver the targeted federal funds rate. This encourages other rates to also fall—from the prime rate to home mortgages and more. This helps to encourage more borrowing and spending. Simultaneously, it discourages saving because the future value of money falls.

Conversely, the Fed can choose to tighten its policy stance. When it wants economic activity to slow down—for example, because it sees rising inflation in the economy—the Fed can raise its target for the federal funds rate. To do so, it will use an open-market operation to sell government bonds to banks. This will reduce the quantity of reserves supplied, which will drive the federal funds rate higher. Once again, the Fed will calibrate the size of the open-market operation to deliver the targeted federal funds rate. This causes the prime rate and other riskier rates to rise. Higher rates discourage spending and borrowing and encourage saving. Finally, if the Fed thinks that current economic conditions are consistent with its mandate, it will adopt a neutral policy stance, in which it aims to leave the federal funds rate unchanged.

## 7. Two Interest Rates to Rule Them All?

Beginning in 2008/2009, the federal funds market changed dramatically. After the 2008/2009 financial crisis, the Fed increased the supply of money sharply to boost the economy and drove the federal funds rate down to zero. Even after this, the Fed continued to engage in large-scale bond purchase programs, in which it purchased securities from banks in exchange for reserves. These additional purchases created an enormous increase in bank reserves, which the banks were able to lend on the federal funds market. This combined a sharp increase in the supply of funds with a sharp decrease in the demand for funds on the federal funds market. Banks were flush with reserves and were happy to lend but were not looking to borrow.

Furthermore, the discount rate was lowered to 0.5%, and the interest rate on excess reserves was lowered to 0.25%. Concerning the federal funds market, the intersection of the supply curve and the demand curve would basically be at the floor determined by the interest rate on reserves. Furthermore, increases in how much funds the Fed supplied to the federal funds market no longer lowered the federal funds rate. Banks no longer needed to trade reserves with each other in the federal funds market to meet the reserve requirements.



Starting in the early 1980s, the volume of funds in the federal funds market rose steadily, from \$100 billion to a peak of \$480 billion in September 2008. However, from 2010, the volume fell sharply. The banks had extraordinarily ample reserves, meaning they had far less need to trade in the federal funds market for additional funds. In short, for more than a decade after 2008/2009, banks operated in a different world from the federal funds market. As inflation rose during the third year of the pandemic and the Fed began to unwind its bond purchase programs, a more conventional federal funds market returned. This means the Fed can raise and lower the federal funds target rate as circumstances warrant by changing the discount rate and using open-market operations to reduce the supply of funds in the market.

The federal funds rate has only a limited effect on longer-term interest rates. Consider a bank that has many clients wanting to take out 30-year mortgages when the federal funds rate is low. The fact that the bank can borrow at low rates on the overnight market will not induce it to charge a low interest rate on a 30-year mortgage since it doesn't know if it can keep borrowing on the overnight market at low rates. A long-term rate, such as for a 30-year mortgage, does not move in lockstep with the federal funds rate. But if you remove the federal funds rate and swap in the US Treasury security rate for a bond lasting 10 years, the pattern tracks 30-year mortgages quite closely.

## Understanding Bonds

When large corporations and governments borrow money, they typically do so by issuing bonds. A bond is a financial instrument that offers a payment of a fixed sum of money at a prespecified future date. It is sold in exchange for a sum of money today. The promised future payment is the face value or par value of the bond. The prespecified future date is the maturity date. Sometimes, the bond also offers a stream of payments (i.e., coupon payments, which are calculated as the face value times a fixed interest rate known as the coupon rate) until maturity.

Note the difference between a standard loan and a bond. In a loan, the borrower specifies how much she wants to borrow up front, and the lender responds by specifying what the future repayments should be. In a bond, the borrower specifies how much she is willing to pay in the future, and the lender specifies how much she is willing to lend up front.

## 7. Two Interest Rates to Rule Them All?

How is the price of a bond determined? To keep things simple, consider zero-coupon bonds. This bond is one where the coupon rate is zero, meaning the bond issuer will make no interest payments, paying only the face value of the bond at maturity. For example, a new \$10,000 zero-coupon 10-year Treasury bond promises to pay \$10,000 at the end of 10 years but offers no other payments in the interim. How much should you pay now for this bond? Remember that buying the bond entitles you to \$10,000 in 10 years' time. Since that money is only received in 10 years' time, it is obviously not worth as much as \$10,000 today.

If you can buy this \$10,000 zero-coupon 10-year bond for below \$10,000, you are effectively earning interest by buying it. Similarly, if the seller is receiving some amount of money less than \$10,000 today in exchange for a promise to pay \$10,000 in the future, then they are paying interest. The interest rate here is called the yield of the bond, or the yield to maturity. The yield of a bond can be thought of as the effective interest rate the bondholder will earn if the bond makes its promised payment. If you buy the bond for \$6,000, then the question becomes: What interest rate would turn \$6,000 into \$10,000 in five years? The yield can be calculated as equaling 5.24%. If you paid \$8,000 for this bond, the yield of the bond would be 2.26%.

As the prices of risk-free US government bonds fluctuate, their yields fluctuate in the opposite direction. These bond yields dominate what happens with many other interest rates in the economy. How much someone is willing to pay for any bond depends on four things: risk, inflation, impatience, and alternatives. More riskiness, higher expected inflation, buyer impatience, and attractive alternative investments lower the price of a bond.

Suppose the yield on a 10-year US government bond is 5%. If you have excellent credit and seek a 10-year student loan, the bank will charge you an interest rate greater than the government bond of 5%, say 6.25%. To the bank, you are a greater risk than the US government in terms of your promise to repay money in the future. The additional 1.25 percentage points are the risk premium. If someone with a lower credit score than you goes to the bank and asks for a similar 10-year student loan, that person may be charged an even higher rate, say 7.5%. Their risk premium of 2.5 percentage points is greater because they are riskier than you. What if the yield on a 10-year US government bond falls to 3%? Assuming a constant risk premium, the bank

## 7. Two Interest Rates to Rule Them All?

will now charge you an interest rate of 4.25% and the riskier borrower an interest rate of 5.5%. In this way, the movement of US government bond yields influences the millions of long-term interest rates in the economy.

Risk premiums can also change, driving interest rates even higher, at least for some borrowers. For example, in the aftermath of the subprime mortgage crisis, borrowing costs for homeowners with moderate credit scores rose because banks wanted to move away from risky home loans.

There are several services that rate the reliability of bonds, such as Moody's. Typically, the higher your bond rating, the lower the interest rate you have to offer to compensate people for holding your bond. Thus, companies with an AAA rating can borrow money more cheaply than those with a CCC rating. These ratings agencies seek to summarize the likelihood of default: how likely it is that a borrower will not honor the promised payment at maturity.

This applies to individual states and municipalities as well. For example, the state of Illinois slipped below other states to a rating of BBB from 2017 to 2020. There are also differences across countries. Bonds issued by the US or German governments have typically been rated AAA, while those issued by Greece might be rated BB.

### READING

Hill, Claire A. "Why Did Rating Agencies Do Such a Bad Job Rating Subprime Securities?" *University of Pittsburgh Law Review* 71 (2009): 585.

Ihrig, Jane, and Scott Wolla. "Teaching about New Monetary Policy Tools." Federal Reserve Bank of St. Louis, August 5, 2021. <https://www.stlouisfed.org/timely-topics/teaching-about-new-monetary-policy-tools>.

PIMCO. "Everything You Need to Know about Bonds." Accessed December 12, 2022. <https://www.pimco.com/en-us/resources/education/everything-you-need-to-know-about-bonds>.

Swanson, Eric T. "The Federal Funds Market, Pre- and Post-2008." Working Paper 29762, National Bureau of Economic Research, 2022. <https://www.nber.org/papers/w29762>.

The Federal Reserve. "Policy Tools." Accessed December 12, 2022. <https://www.federalreserve.gov/monetarypolicy/policytools.htm>.

# 8

## WHY ECONOMIC GROWTH IS SO POWERFUL— AND HARD

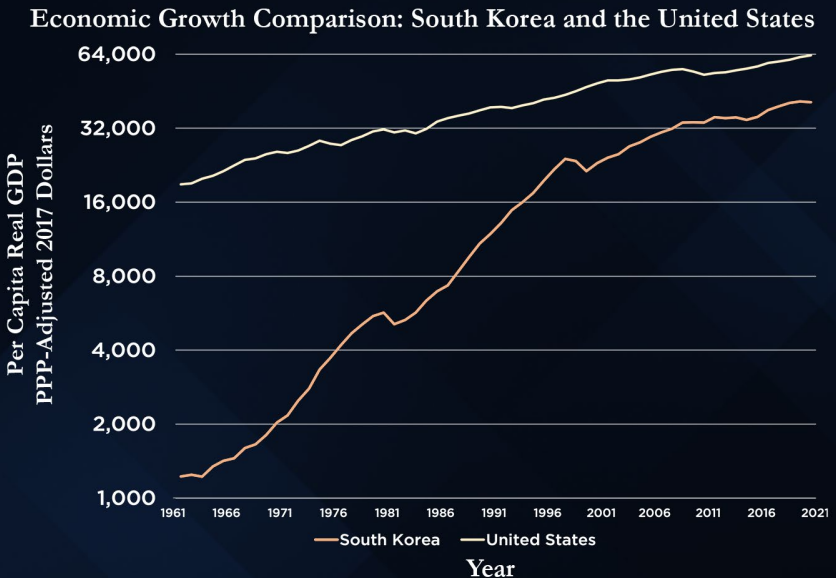
**F**our decades of rapid economic growth in China have transformed the lives of 1.4 billion people. If countries in South Asia and sub-Saharan Africa can also achieve sustained economic growth over the next generation, much of the abject poverty and misery that remains in the world will be erased. This lecture will explore why sustained economic growth is important and why it is so hard to achieve. In particular, it will introduce you to the work of Robert Solow, who developed an influential model of economic growth. He showed macroeconomists that sustained economic growth is not an inevitable, or even a natural, feature of an economy.



## What Is Economic Growth?

Economic growth essentially refers to long-run movements rather than short-run fluctuations in per capita real GDP. In other words, faster economic growth means a steeper long-term trend in real GDP per capita. Most economists use a log-scale diagram when looking at data on growth. On such a diagram, a doubling from \$20,000 to \$40,000 in per capita income, for example, has the same vertical gap as a doubling from \$40,000 to \$80,000.

Different countries have different growth paths. For instance, considering PPP-adjusted real GDP per capita data from the International Comparisons Project, the growth path for the South Korean economy is much steeper than that for the US economy. Over six decades, starting in 1960, South Korea's real per capita GDP doubled five times: from a little more than \$1,000 to \$32,000. The United States' per capita GDP was 16 times higher than South Korea's in 1960 but only 1.6 times higher in 2021. In essence, South Korea has closed



90% of the gap between the two nations in 60 years. It achieved that doubling by growing steadily at a rate of around 6% per year versus 2% per year for the United States.

Small differences in growth rates, sustained over an extended period of time, can have significant implications for living standards. A quick estimate is possible using the rule of 70. This rule of thumb says that if an economy has a growth rate of  $g\%$  a year, then it doubles roughly every  $70/g$  years. For example, if an economy grows at 1% a year, it will double every 70 years. South Korea grew at an average of 6% a year, meaning per capita real GDP doubled at just under every 11.7 years. Over the 60-year period, per capita real GDP doubled at least five times. Doubling five times,  $2 \times 2 \times 2 \times 2 \times 2$ , means a 32-fold increase in South Korea's GDP. Had it grown at 2% a year—as the United States did over this period—it would have doubled per capita income every 35 years. At this rate over a 60-year period, income would have come close to doubling twice: a fourfold increase.

## Benefits of Rapid Economic Growth

The rule of 70 shows that even a 1- to 2-percentage-point change in the rate of growth, when sustained over a long period of time, can lead to large changes in income. If you know what drives growth at 4% or 6% instead of at 2%, then you can make significant improvements in living standards worldwide. For instance, per capita real GDP is strongly correlated with reduced infant mortality, or the number of deaths before age five per 1,000 live births. There is a wide disparity globally. In poor countries such as the Central African Republic or Somalia, infant mortality is between 100 and 200 deaths per 1,000 births. Meanwhile, in rich countries such as Singapore or Norway, infant mortality is around 2 deaths per 1,000 live births. Substantial reductions in infant mortality seem to be achieved at around \$8,000 of per capita GDP, some more by \$10,000, and virtually all by \$20,000.

In general, development indicators rise with per capita GDP. The largest gains come when countries move from per capita GDP levels of less than \$1,000 to per capita GDP levels in the \$5,000 range. Such benefits far exceed what can be achieved with a jump from a \$10,000 per capita GDP to a \$50,000 per capita GDP economy. The latter jump is more likely to be noticed in material wealth improvements and quality of life improvements. Achieving

fundamental improvements requires increasing a country's growth rate by a couple of percentage points a year and sustaining it for a generation or two. If you can sustain 4% growth for a 35-year period, per capita income will double two times. Two rounds of doubling is enough to turn a \$500 per capita GDP economy into a \$2,000 per capita GDP economy.

Why are some countries able to grow while others are not? Consider all the countries in the world that had a per capita real GDP of less than \$2,000 in 1994 and compare how fast they grew (on average) during the period from 1994 to 2019. There is clearly a divergence of experiences. Some low-income countries, such as Equatorial Guinea, grew especially fast, at a rate of more than 8% a year. This would mean that their per capita real GDP would have doubled every 8.75 years, which means close to three doublings (an eightfold increase) over this period of 25 years. Some other countries, such as Burundi, experienced a negative real GDP growth rate over this period. Per capita real GDP is lower than it was a quarter century ago.

Even high GDP per capita does not tell you anything about the distribution of income. For example, what about Equatorial Guinea's extremely fast growth rate of more than 10% each year? That was driven by the discovery of offshore oil deposits. However, leaders of the country expropriated most of the revenues generated by the oil reserves. Most of the roughly 1.5 million people living there inhabit a country that remains one of the most poverty-stricken and unequal nations in the world. Thus, achieving per capita economic growth is a necessary, but not a sufficient, condition for improving the well-being of most people in a country.

## Solow's Model

To uncover the challenges of delivering sustained economic growth, macroeconomists use the most famous model of economic growth pioneered by Robert Solow. In its simplest form, the Solow model specifies a production function. This is a mathematical relationship that describes how inputs to production (capital, labor, education, or resources) are translated into output. In Solow's model, the production function has three important properties or assumptions. The first assumption is that inputs are useful: If you add more of an input, you will get more output.

## 8. Why Economic Growth Is So Powerful—and Hard

The second assumption is that there are diminishing returns to adding more of any single input. The idea here is that inputs work in conjunction with other inputs. If you add more of one input, holding other inputs constant, you get less and less additional production from that input. For example, given one plot of land and one farmer, moving from zero tractors to one tractor increases output more than moving from one tractor to two tractors. Intuitively, without additional land or labor, the second tractor is less useful than the first.

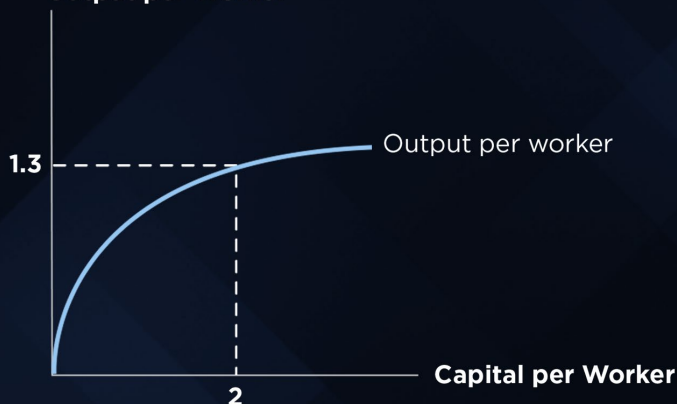
The third assumption concerns constant returns to scale; that is, if you scaled all inputs by some factor  $x$ , then you would increase output by the same factor  $x$ . In other words, increasing all inputs by 5% would increase output by 5%. This is a scaling argument: Two farmers, two tractors, and two plots of land will produce twice as much output as one farmer, one tractor, and one plot of land.

The amount of output that the economy produces rises with capital input per worker. However, each additional unit of capital input per worker brings less and less additional output. If you double the amount of capital and the number of workers, you get twice as much output. For example, suppose there were 10,000 units of capital and 5,000 workers in the economy, producing

Model Economy:

Capital = 10,000 Workers = 5,000 Output = 6,500

**Output per Worker**



## 8. Why Economic Growth Is So Powerful—and Hard

6,500 units of output. Then, capital per worker would be 2, and output per worker would be 1.3. However, if you scaled everything up by a factor of 2, you would have 20,000 units of capital and 10,000 workers. Thus, capital per worker would still be 2. Constant returns to scale would tell you that you would produce twice as much output as before: 13,000 units. Thus, output per worker will remain at 1.3. The capital per worker and output per worker would be unchanged—in other words, the economy's position would not change.

With this understanding of the production process, Solow made two more key simplifying assumptions: First, he assumed that the economy reinvests a constant fraction of its output into capital. Next, he showed that a certain amount of investment is needed to accommodate growth in the labor force and to replace depreciated capital in the economy. For simplicity, Solow assumed that both the labor force growth rate and the depreciation rate would be independent of a country's per capita GDP level. Relaxing this assumption complicates the model but does not change its key insights. The more capital per worker an economy has, the more it needs to invest to maintain the depreciating capital it already has.

What Solow's model shows is that the economy moves over time to a level of capital per worker that is known as its steady state. That point is the intersection of how much investment the economy is doing with how much investment it needs to sustain its current level. Why does this work? To the left of the intersection point, the economy is investing more than the required amount. Since it is adding more capital than needed to maintain existing levels, the amount of capital per worker will increase over time. However, because of diminishing returns, the gap between how much is invested and how much needs to be invested will gradually decrease until they are exactly equal at the steady state. When that point is reached, it tends to stay there.

Conversely, to the right of the intersection point, the economy is investing less than the required amount: Since it is adding less capital than needed to maintain existing levels, the amount of capital per worker will decrease over time until the steady state is reached. Over time, no matter where it starts, the economy ends up at the steady state. This is essentially the gist of Solow's model. In its more complicated form, it shows that several factors contribute

to a higher steady state: high investment rates in capital, high levels of education, high labor force participation rates, and an abundance of accessible natural resources.

The model also shows that low population growth rates also contribute to a higher steady state. The basic intuition is rooted in the concept of diminishing returns. Higher rates of population growth mean that the economy has more workers but also more mouths to feed. With diminishing returns, a poor country that is unable to increase its capital stock will find that each additional worker contributes less and less to additional output. Thus, steady-state output per worker will decrease. However, there are three more critical insights that are much more revealing.

### Solow's Critical Insights

The first key result of Solow's model is that the growth rate of steady-state per capita real GDP in this economy is zero. When the economy reaches its steady state, it stays there. In other words, there is no natural tendency for an economy to deliver sustained economic growth. Why? Imagine building 400 miles of true high-speed rail (HSR) lines connecting Washington DC, Philadelphia, New York, and Boston. This could cut travel time between Boston and DC from the current eight hours to two hours and dramatically increase economic activity along a key corridor in the United States. By contrast, in China, which already has an extensive HSR network, adding 400 more miles would do much less to spur additional economic activity. Thus, even though you can increase GDP by adding more inputs, each additional input is less and less productive on its own.

Constant returns tells economists that to scale an economy up, all of the inputs need to be scaled up. However, many inputs—natural resources, land, working-age population—run into natural constraints that make them hard to scale up. Sustaining economic growth will become harder as more inputs are added. In short, there is no natural tendency for an economy to deliver sustained economic growth.

The second key result of Solow's model is about moving to a higher steady state. In particular, if a country can increase its steady state, it will grow faster and move toward that new steady state. The more it can push its steady state

out, the faster it will begin to grow. Consider the economic resurgence of China since 1978. China's economic stagnation in the 1950s and 1960s had led to the deaths of tens of millions of Chinese. The most populous country in the world was also one of the poorest at the time that Deng Xiaoping took over in 1978.

Deng instigated a series of reforms that brought millions of Chinese workers into the formal labor sector and away from peasant farming. He invested in factories that manufactured simple goods and encouraged foreign firms to enter into specially designated export zones where they could employ China's abundant cheap labor to produce goods for export. His draconian one-child policy cut back on population growth. Simultaneously, he improved education opportunities for young Chinese and encouraged many to go abroad and acquire educational skills that they could then bring back to China as doctors, engineers, and scientists. These changes increased the amount of investment that China was doing (as a result of higher domestic and foreign investment). They also decreased the amount of required investment (because of lower population growth and a more skilled workforce that was capable of maintaining existing capital).

However, note that Solow's model shows that if China wants to keep that economic growth going, it will have to keep pushing its steady state out. It will not be easily able to do so by accumulating more of the same inputs. When deciding whether to build new infrastructure, policymakers should weigh the additional economic output it produces against its maintenance cost. If the output outweighs the cost, a country should keep investing. However, because of diminishing returns, the additional benefit of each new investment is expected to dwindle. At some point, the benefits will no longer outweigh the costs, and growth will slow down as it approaches steady state.

In the early 1990s, Alwyn Young did a detailed study of the East Asian "tiger" economies that grew rapidly in the 1970s and 1980s. Young showed that these East Asian countries made remarkable progress in labor force participation, educational attainment, and investment in physical capital during these decades. Between 1965 and 1990, these changes raised their steady states and created room for these economies to grow. However, Young cautioned that as these countries approached their new steady states, their economic growth

### Economic Growth and Input Growth in Selected East Asian Countries (1965–1990)

	Singapore	South Korea	Taiwan	Hong Kong
GDP growth	8.50%	8.50%	8.60%	7.30%
GDP per capita growth	6.60%	6.90%	6.80%	5.70%
Labor force participation	27% → 51%	27% → 36%	28% → 37%	38% → 49%
Investment/GDP	10% → 41%	8% → 40%	9% → 21%	30% → 30%
No schooling	55% → 0%	31% → 6%	17% → 4%	19% → 6%
Primary school only	28% → 34%	42% → 19%	57% → 28%	54% → 23%
Secondary school	16% → 66%	27% → 75%	26% → 68%	27% → 71%

would slow down, as Solow had predicted. He was pointing out, for example, that it would not be possible for Singapore to quadruple its investment-to-GDP rate from 40% to 160%, even though it had quadrupled from 10% to 40%.

Young's warnings were a prescient warning that past growth rates cannot last forever if they are driven purely by growth in the inputs that have already brought them so far. Note that Solow's key result is not saying that economic growth via input accumulation is impossible. A country that invests more, finds more natural resources, better educates its people, or acquires more skilled immigrants will increase its steady-state real GDP per capita. Such an economy will grow from its current level to reach that new, higher steady state. However, it would become harder to push that steady state higher by adding more inputs.

This leads to Solow's third key result, which explains how countries continue to grow for sustained periods of time. By showing the limits to which an economy can achieve sustained economic growth based only on input accumulation, Solow showed that making better use of, or improving, existing inputs has to be the driver of sustained economic growth. Technology is the engine of economic growth. When looking over the half century from 1970

to 2020, economists can declare that never before in human history have the living standards of so many people improved in such a short period of time. This transformation was made possible by sustained economic growth.

Yet Solow's influential work showed that there is no natural tendency for an economy to experience sustained growth in per capita real GDP. In fact, in the absence of technological change, the long-run natural tendency of an economy is to not grow at all. Countries can and should grow faster for a while—even for a generation or two—by investing more, bringing more people into the labor force, or educating more workers. However, diminishing returns mean that such input-driven growth alone will eventually be expected to slow down. Sustaining growth requires understanding the process of technological progress.

## READING

Collier, Paul. *The Bottom Billion: Why the Poorest Countries Are Failing and What Can Be Done about It*. Oxford University Press, 2008.

Dizikes, Peter. "The Productive Career of Robert Solow." *MIT Technology Review*, December 27, 2019. <https://www.technologyreview.com/2019/12/27/131259/the-productive-career-of-robert-solow/>.

Krugman, Paul. "The Myth of Asia's Miracle." *Foreign Affairs* (1994): 62–78.

Panagariya, Arvind. *New India: Reclaiming the Lost Glory*. Oxford University Press, 2020.

Solow, Robert M. "A Contribution to the Theory of Economic Growth." *The Quarterly Journal of Economics* 70, no. 1 (1956): 65–94.

Vogel, Ezra F. *Deng Xiaoping and the Transformation of China*. Belknap Press of Harvard University Press, 2011.

# 9

## TECHNOLOGY: THE ENGINE OF SUSTAINED GROWTH

**R**ecall that the constraints imposed by diminishing returns, coupled with resource constraints and finite labor forces, mean that sustained economic growth will slow down over time. Yet some countries exhibit sustained economic growth that continues for more than a single generation. By extending his theoretical work to incorporate technological progress, Robert Solow concluded that sustained economic growth is only possible via sustained growth in technology. This lecture will explore what economists mean when they talk about technology, why technology is different from other inputs, and the challenges an economy faces in trying to incentivize greater technological progress.



## What Is Technology?

Economists define *technology* broadly as the ideas, or knowledge, that help people produce output from inputs. Having more technology means being able to produce more output from a given amount of inputs. Thus, technology is a broad concept that can encompass various discoveries, inventions, and innovations, including scientific discoveries and efficiency improvements that come from specialization. Here, technology is the idea behind the machine—for example, how you build a faster computer—rather than the machine itself.

Technology has an important role to play in overcoming the limitations imposed by diminishing returns to labor and capital. Societies have often learned to produce more with less of the scarce inputs. Technological progress is fundamentally different from accumulation of inputs such as land, raw materials, or number of people. First, ideas are non-rival by nature: The use of an idea by one does not preclude its simultaneous use by another. In short, an idea can be used by millions of people at the same time.

Second, ideas are also inherently non-excludable by nature: It becomes possible to exclude others from using an idea only when there are legal constraints that limit the use of the idea to those who have paid for the right to use it. These constraints are called intellectual property rights. They include patents, copyrights, and trademarks. For example, a pharmaceutical company is not allowed to reverse engineer Moderna's COVID-19 vaccine and charge people for its use.

The non-rival nature of ideas implies that technology is different from regular inputs, which are characterized by diminishing returns. Think back to the production function. If one farmer, one tractor, and one plot of land could produce five tons of a crop, adding a second tractor would not lead to as much additional output as adding the first tractor did. This is because the tractor is a rival good. When the farmer is using the first tractor, he can't use the second tractor. Moreover, if one farmer is planting crops in that plot of land using a tractor, then another farmer cannot cultivate his crops using his tractor on that same piece of land. This leads to the concept of constant returns to scale. You need two farmers, two tractors, and two plots of land to produce 10 tons of a crop.

Suppose you add technology to the production function. The output depends not only on farmers, tractors, and land but also on ideas—for example, what crops grow best in that type of soil. Now, one farmer, one tractor, and one plot of land can produce more than five tons of the crop, armed with the knowledge of what crops grow best in that type of soil. More importantly, the idea behind a technology is non-rival. Thus, to double the output, both farmers can use the same idea without hindering each other. Suppose there is a second idea: the optimal watering cycle to maximize crop output. There is no reason to believe that the second idea will contribute less to additional output than the first idea did. The ideas could reinforce each other, meaning that the second idea is more valuable in the presence of the first idea.

## Measuring Technological Progress

Solow incorporated technology into his canonical model of growth and showed that the steady-state per capita real GDP level that the economy would converge to would no longer be a fixed value. Instead, the economy would converge to a sustained growth path where real GDP per capita would grow at a constant rate over time. More importantly, the rate of growth of per capita real GDP over time would be the rate of the growth of technology. In other words, in the long run, economic growth would be driven entirely by technological progress.

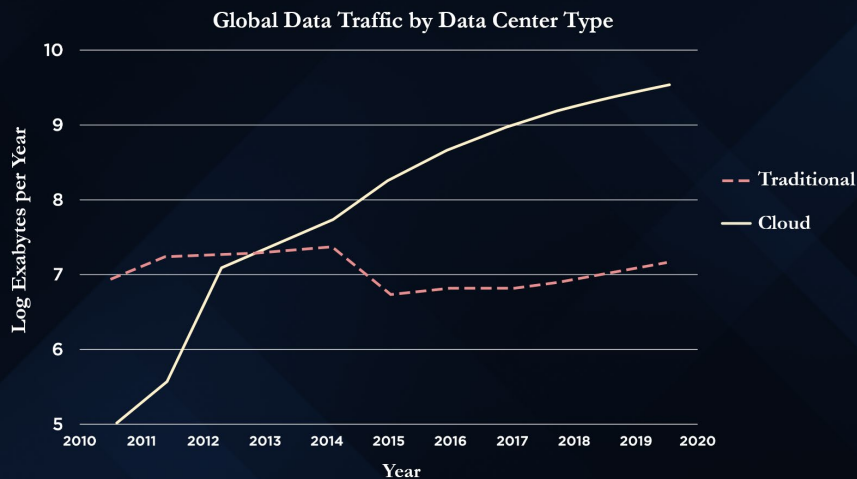
Suppose you were the leader of a country who had learned that the long-run growth path of your economy depended on how fast your country was able to add to its stock of productive ideas. How would you know how fast technology is growing in your economy? Solow laid out a series of steps that could be used to measure the growth rate of technology: First, specify a production function that shows how inputs come together to produce per capita output. Second, accurately measure how much the inputs grew. Third, use the production function to calculate how much output would have grown as a result of those input increases. Fourth, calculate the difference between how much output grew and how much calculated output would have grown using the production function. This difference is the output growth that is driven by technological progress.

Economists call this difference the growth in total factor productivity (TFP): a measure of how efficient the economy is at producing output from a given quantity of inputs. In short, Solow showed economists that the way to measure TFP growth is as a residual. Technological progress is the portion of economic growth that cannot be explained by growth in inputs. If output grows by more than you would expect it to have grown by the addition of inputs, then, by definition, it must have grown because of better use of the inputs. This difference is often called the Solow residual.

## Calculating the Solow Residual

Due to measurement errors and other problems, many calculations need to be done correctly to arrive at an informative Solow residual. Consider Daniel Sichel's work on cloud computing as an example of how to accurately measure technological progress. In recent years, many companies have started outsourcing their computing needs to cloud computing services such as Amazon Web Services. If you were unaware that cloud computing was in use, you would observe the company reducing its inputs—buying fewer computers, servers, routers, and networking gear and reducing its employment of network engineers, technicians, and programmers. However, it seems to be producing just as much, if not more, output.

In his paper, Sichel and his coauthors show how fast cloud computing centers grew in terms of handling data traffic during the 2010s when compared to traditional workplace-based computing. If you naively applied Solow's method and measured only traditional computing, you would conclude that there was technological progress: Companies are producing as much output with fewer computing inputs. However, this is an inaccurate conclusion. The company has not cut back on its use of computer services. It has simply shifted to using cloud computing services. Thus, researchers like Sichel have to come up with ways to quantify the use of cloud computing services in the economy. Accounting for the increased use of cloud computing services will reduce the incorrect estimate of technological progress in the economy.



However, note that it is still possible that cloud computing has resulted in technological progress. If Amazon Web Services can provide more powerful computer resources that do faster or more reliable calculations or reduce computing downtime, the company may be able to produce more output with less spending on computing resources. Sichel's research shows that between 2009 and 2016, the price of cloud data storage fell by 17% a year, and the price of cloud computing fell by 7% a year. Companies can do more with less thanks to cloud computing.

Careful measurement of technological progress is also helpful in developing a proper perspective on growth miracles like China. In the first quarter century after Deng Xiaoping's economic liberalizations (between 1978 and 2003), per capita GDP in China grew at around 8% a year. You know from the rule of 70 that per capita GDP would have doubled every 8.5 years. Over a 25-year period, it would have doubled three times. Therefore, China's 2003 per capita GDP was eight times as large as it was in 1978. You also know that growth would slow down to match the growth rate of technology as it approached steady state.

If China were to continue on that path for the next 75 years, then by 2078, its economy would have doubled another nine times. However,  $2$  to the ninth power equals  $512$ , meaning you would be saying that real per capita GDP in 2078 in China will be 512 times what real per capita GDP was in 2003, when it was around \$5,000. That would be a real GDP of \$2.5 million per person. Using Solow's residual to calculate how fast technology grew in China during this post-reform period would help you understand what a more sustainable long-run growth rate might be.

In 2008, economists Barry Bosworth and Susan Collins examined the success stories of China, India, and East Asia. They looked at how much GDP growth increased in each region and then calculated how fast labor, capital, and education grew in these countries. They then used Solow's method to calculate the growth rate of TFP. For example, between 1978 and 2004, China's GDP per worker grew at 7.3% a year. In total, 3.2 percentage points of that growth came from adding capital inputs and 0.3 percentage points came from educational improvements. The authors concluded that 3.6 percentage points of growth came from technological progress. Based on Solow's results, China's long-run economic growth would be much closer to 3.6% than to 8%.

A few things to note: First, this does not mean that technology-driven growth is the only thing that matters. China was clearly better off by growing at 8% a year rather than 3.6% a year. The former was only possible because of all the progress China made in terms of economic reforms that encouraged more inputs into production. Second, the measure of 3.6% technological growth reflects a measure of ignorance about China. If the authors mismeasured the extent by which inputs increased in China, they would also miscalculate the extent of technological progress. Finally, even if they did accurately measure technological progress in China between 1978 and 2003, there is no guarantee that China could sustain that progress.

# Improving the Rate of Technological Progress

In the late 1980s and early 1990s, Paul Romer provided insights into how new technologies are created in the economy. His work focused on the role of economic incentives and institutions in driving those creations. First, sometimes, a new technology is a happy accident or serendipity from another activity. For example, a pharmaceutical product created for one purpose may be found during clinical trials to be useful for a different clinical purpose. Second, new technology comes from learning by doing. For instance, someone doing a rote task determines an alternative way to do the task. Third, technology can be the result of a systematic search to find new and better ways of doing things. Research by Tesla to develop an electric car is one such example.

Romer highlighted the importance of the non-excludable and non-rival nature of ideas. Because ideas are non-rival, many people can use the idea at the same time. However, insofar as the idea is also non-excludable, the people who use the idea can't be charged for its use by the creator of the idea. Romer pointed out that an important feature of ideas is that they take more resources to create than to use. Moderna would have to invest a couple billion dollars in developing the vaccine and carrying out efficacy and safety tests. However, once the vaccine is released, it would be much more straightforward for a rival pharmaceutical company to copy the vaccine. The combination of high fixed costs and low marginal costs, combined with non-excludability, would result in people being reluctant to invest the upfront money necessary to create an idea. Without a subsidy or a system of property rights to the idea, researchers will not develop the idea because after they incur the fixed cost, others can enter the market and charge much lower prices for their goods.

To entice researchers to come up with ideas, especially ideas that are not excludable, Romer argued that policymakers should devise methods to subsidize fixed costs (grants from the government) and set up intellectual property rights (patents, copyrights) that restrict the ability of anyone other than the holder of the property right to reproduce it. This monopoly power provides incentive for people to engage in the research effort needed to create the idea. Without the intellectual property right, the idea may never be created. What happens when

## 9. Technology: The Engine of Sustained Growth

intellectual property rights are used to make an idea more excludable? This effectively creates artificial monopoly power for the creator—for example, only Moderna is able to sell its mRNA vaccine to the government.

Note that there is always a trade-off between incentive and access when it comes to technology. Providing a patent for a new invention ensures that the creator is rewarded for their efforts and serves as an incentive for others to come up with a discovery. However, since others must pay the creator to use that technology, access to that technology is limited. The key for a country is to find the right balance between encouraging innovation and ensuring access to the benefits that flow from those innovations.

For example, suppose you were the head of the National Institutes of Health in January 2020 when the first signs of COVID-19 began to appear. You are tasked with devising the most effective way to incentivize the creation of a vaccine that is widely accessible to the population. You are given a budget of \$100 billion. Here are the possible paths you could choose. In path A, you can distribute the \$100 billion in research grants among all firms that engage in vaccine research and development. As a condition of receiving funding, no patent will be awarded to any firm that receives a grant. In path B, you can award cash prizes to the first five firms that come up with a new vaccine, with the largest prize given to the first firm that comes up with a successful vaccine. Again, a condition of receiving the cash prize will be that the invention does not get a patent. In path C, you can allow any company that develops an effective vaccine to get a patent and instead use the \$100 billion to provide subsidies to those who can't afford vaccines.

**Path A:** Distribute \$100 billion in R&D grants to firms, no patents

- Any firm that does research can apply for grant
- If you receive a grant, no patent protection on vaccine

**Path B:** Award \$100 billion in cash prizes to firms, no patents

- First to market gets bigger share of prize
- If you receive a prize, no patent protection on vaccine

**Path C:** Use \$100 billion for subsidies, allow patents

- No firms receive funds, only consumers
- If you develop a vaccine, you get patent protection

There are pros and cons of each approach. First, consider path A: research grants with no patent. If a vaccine is developed, patients will be able to have access to this vaccine without paying high prices. However, the grants may be insufficient to cover the research costs, particularly if distributed among many firms. If 100 firms each get \$1 billion from the government, it may not be enough incentive for them to do research to create a vaccine. Next, consider path B: the “prize but no patent” idea. This also has the attraction of enhanced access to a successful vaccine. However, rewarding speed over efficacy may mean that firms may rush a less-safe vaccine to be first to market. Finally, path C uses patents as the incentive for companies to come up with new vaccines. The \$100 billion enhances access to vaccines for people. However, the company may charge a high price for the vaccine, knowing that people are eager to get a shot to protect themselves from the virus and that money has been earmarked for vaccine purchases. Access to the vaccine may end up being limited despite the subsidies.

Policymakers who are responsible for intellectual property rights face a continuous battle between incentive and access. If intellectual property rights are too weak, invention and idea creation are discouraged. However, if the rights are too strong, access to the multitude of new ideas is limited.

## READING

Bosworth, Barry, and Susan M. Collins. “Accounting for Growth: Comparing China and India.” *Journal of Economic Perspectives* 22, no. 1 (2008): 45–66.

Clark, Gregory. “Winter Is Coming: Robert Gordon and the Future of Economic Growth.” *American Economic Review* 106, no. 5 (2016): 68–71.

Gordon, Robert J. *The Rise and Fall of American Growth*. Princeton University Press, 2016.

Kugler, Maurice. “The Economics of Ideas: Paul Romer Receives the Nobel Prize.” Berkeley Economics, October 17, 2018. <https://www.econ.berkeley.edu/content/guest-post-economist-ideas-paul-romer-former-berkeley-economics-professor-receives-2018>.

Posner, Richard A. “Intellectual Property: The Law and Economics Approach.” *Journal of Economic Perspectives* 19, no. 2 (2005): 57–73.

# 10

## WHAT ARE FUNDAMENTAL BARRIERS TO GROWTH?

**E**conomic growth can be spurred on by adding more inputs and sustained by technological progress. However, many countries are still struggling to achieve economic growth. What factors limit a country's ability to add inputs or improve technology? This lecture will look at four fundamental factors—openness, conflict, geography, and institutions—that have shaped the economic fortunes of countries worldwide. You will look at how each factor can impact economic growth by affecting the amount of capital, labor, or technology an economy has.



## Openness and Economic Growth

Here, openness is the extent to which domestic economic activity can be influenced by non-domestic entities. It can thus reflect the movement of goods and services, people, ideas and knowledge, or money. On the financial side, openness creates an important wedge between the amount of domestic saving and the amount of domestic investment. In a closed economy (i.e., one that does not allow flows of capital to and from the outside world), domestic saving has to equal domestic investment. Suppose you were living in a closed economy and wanted to invest \$10 million in a palm oil refinery. Your investment has to be financed by domestic savings—either your own or from a loan made by a bank (which would be lending out someone else's savings). Thus, to have growth, you need more investment. However, more investment requires more saving. To have more saving, you need to have more income, which requires more growth and, thus, more investment.

This constraint on saving remains a drag on many poor countries. It's hard to find the resources to fund investments such as better schools or roads. In an open economy, domestic investment does not have to be constrained by domestic saving. Investors have access to foreign capital, which makes it more likely that they can find financing for their projects. In other words, if your palm oil business were a profitable investment opportunity, then you could borrow \$10 million from a foreign bank and use the money to buy equipment. If your economy were even more open, foreigners could build factories and set up businesses of their own accord in your economy. This is known as foreign direct investment.

Access to foreign investment has been an important driving force in many countries. In the post-1978 economic reforms that transformed China's economy, an important role was played by the creation of special economic zones such as Shenzhen and Shantou across from Hong Kong, Zhuhai opposite Macau, and Xiamen on the straits of Taiwan. Those were the earliest zones that encouraged foreign firms to open manufacturing plants that produced goods for export using Chinese labor, foreign funds, and a mix of imported inputs and Chinese-sourced intermediate inputs. Given how poor China was in 1978 and the low capital base it was beginning from, this switch to greater openness was essential for the rapid increase in investment that followed.

## 10. What Are Fundamental Barriers to Growth?

However, the savings of domestic individuals do not have to be invested in the domestic country. If more domestic investment is taken out of the country and invested abroad than foreign investment enters the economy, this could lower investment in the domestic economy. This is why countries can and do choose to be both open and closed along different dimensions. For example, China, from the outset of its post-1978 reforms, was receptive to foreign direct investment in minority shares of companies but much more reticent about allowing foreigners to buy controlling stakes in companies or real estate. Even ordinary investment in Chinese company stocks was highly constrained. China also exerted tight control over the outward flow of money, even in the 2010s. Chinese consumers were forced to keep their money in Chinese-owned banks. This ensured that the Chinese government (and favored Chinese firms) always had access to the large and growing pool of Chinese savings.

## Migration and Outsourcing

Openness can also have an impact on economic growth through the labor force via migration and outsourcing. Immigration can allow a country to attract workers from other countries to make up for shortfalls in the supply of workers available in the domestic economy. For example, the technology boom in the late 1990s in the United States resulted in a large demand for skilled engineers, programmers, and technicians with computer skills. The US education system alone was unable to churn out enough skilled workers to staff those jobs. The shortfall was mitigated by the hiring of skilled workers from foreign countries, including India, China, and the former Soviet Union. This large pool of skilled workers enabled many firms in Silicon Valley to develop new software technologies and computer and semiconductor manufacturers to continue to develop faster processors and computers.

Countries such as the United Kingdom, New Zealand, Australia, and Canada have used priority points systems that award work privileges to young, educated individuals from Asia and Africa to take up jobs as doctors, nurses, and engineers. In the absence of these immigrant workers, economic growth in the receiving country would be severely disrupted. Moreover, the United States and European countries benefit from having workers from regions such as Mexico or South Asia willing to take on arduous and low-paying work in areas such as agriculture and the service sector.

## 10. What Are Fundamental Barriers to Growth?

However, outward migration could be a negative aspect of openness from the macroeconomic perspective of poorer countries. There is the possibility of “brain drain”: the loss of skilled workers to other countries. John Gibson and David McKenzie surveyed the literature and provided estimates of lost GDP around \$4,000 per Indian migrant, \$500 to \$1,000 per migrant from South Pacific islands, and \$6,000 per Ghanaian migrant. However, what might be good for the country may not necessarily be good for the individual or their family. Gibson and McKenzie have shown that the gains for an individual migrant worker from countries such as Ghana are in the range of \$40,000 to \$75,000—an order of magnitude larger than the cost to the sending country.

Remittances are another vital benefit of outward migration. Migrant workers send money back to families in their home countries. These funds are a vital source of funding for building homes or educating family members. In Mexico, remittances account for 4% of GDP; they are higher than 6% for Vietnam and almost 10% for the Philippines. Furthermore, high-skilled migrant workers from poor countries are the ones who are most likely to remit, according to Gibson and McKenzie. Keep in mind also that there are ties that bind. For example, countries such as Korea, Ireland, and India, which have seen large amounts of skilled labor emigrate in the past, have benefited from return investments. Those who have been successful in other countries return to start businesses in their homelands. Finally, the incentive to migrate may lead to greater investments in education that have positive domino effects even on those who do not migrate. Overall, outward migration could also be seen as a positive factor driving growth.

This leads to a second way that richer countries can make up for shortfalls in domestic labor: setting up factories and workplaces in other countries. If bringing workers to the rich country becomes a challenge because of inflexible immigration laws or



## High-Skilled Migrants from Poor Countries Are More Likely to Remit



hostility to immigrants, the company can take jobs to the workers. That's called outsourcing. For example, a software manufacturer facing a shortfall in the number of US programmers may set up a subsidiary in India.

Openness and technology can be linked. Creating the right environment requires financial resources for research, skilled scientists and engineers, and a strong system of intellectual property rights that allow incentives for firms who develop new technologies. Many developing countries find it difficult to develop these new technologies by themselves. Thus, they have to look to foreign countries as sources for acquiring technology. Some of the most significant medical interventions of recent times—the treatments for Guinea worm and river blindness—have been triumphs of openness. Replicating

those successes with endemic diseases such as malaria or dengue fever will have dramatic benefits for economic growth in some of the poorest countries in the world.

# Conflict

A fundamental factor that disrupts growth is conflict. It begins with the substantial economic costs of conflict borne by individuals and families, particularly in the form of loss of life of young men, who are the primary combatants. Even when people are able to flee the conflict zone, millions end up in underfunded refugee camps—often located in equally poor neighboring nations—with the attendant higher incidence of disease and infant mortality. Long-lasting conflict drains resources from education, leaving the next generation less equipped or able to take advantage of even a brief window of opportunity. The already small segments of the population who are well educated or have the wealth to fund entrepreneurial activities flee the country in search of a better place to raise a family.

Conflict also results in the destruction of capital. Roads, bridges, and power lines are often destroyed directly or as collateral damage. Because money borrowed for soldiers, tanks, guns, and bullets in wartime must be paid back in peacetime, that means less infrastructure can be built even after the conflict is over. Moreover, long-term instability makes people reluctant to enter into long-term contracts. Thus, borrowing money to start a business becomes challenging. Uncertainty about property rights leads to a greater reluctance to make long-term investments. Foreigners become reluctant about investing in countries racked by turmoil, where their investments could be at risk of destruction or appropriation. Politicians become more rapacious and focused on maximizing their own personal wealth instead of building more stable institutions.

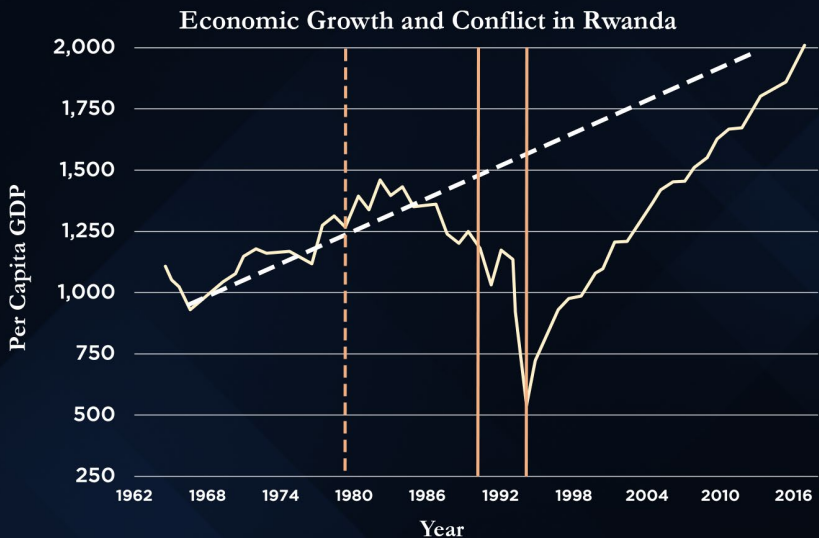
Finally, conflict is also not conducive to most forms of technological progress. A focus on military technology to the exclusion of civilian technology, disruptions to education, capital flight, and lack of foreign investment in civilian technology take a toll on technological progress. Time, patience, and a belief in better times in the future are scarce commodities during times of conflict.

## 10. What Are Fundamental Barriers to Growth?

A recent paper by Natalija Novta and Evgenia Pugacheva in the *Journal of Macroeconomics* calculated that the macroeconomic impact of a conflict typically results in per capita GDP being 28% lower than the counterfactual a decade after the onset of the conflict. Brock Blomberg and Greg Hess have shown that the economic impact of conflict is akin to a 30% tariff on trade in terms of reduced economic activity between countries. Consider the experience of Rwanda since it gained independence in 1962. The horrific civil war in 1994 cut per capita real GDP by half in only five years. More than 25 years of sustained growth since the end of the conflict have made Rwanda an African success story in many people's eyes. However, essentially, Rwanda has only managed to eke back to where it would have been if you extended out the pre-conflict counterfactual growth path.

# Geography

A third fundamental factor that can impact a country's ability to achieve economic growth is geography. In the early 2000s, Professor Jeffrey Sachs of Columbia University highlighted the vast disparity in per capita real GDP



## 10. What Are Fundamental Barriers to Growth?

between tropical and temperate countries and sought to find geography-based explanations for this disparity. For example, many of the world's deadliest diseases, such as malaria, are rampant in tropical climates because the agents of transmission (e.g., mosquitoes) thrive there. These tropical diseases have taken an unimaginable toll in terms of productive years of life lost.

Another way geography has impacted growth is related to access to the ocean. Landlocked countries and countries without navigable rivers have a much greater challenge because it is more difficult to move goods across land than water. Moreover, lack of access to a port leads to difficulty integrating into the global economy. Building roads to transport goods in a mountainous, landlocked nation such as Nepal is a much greater challenge than in a nation with a deepwater port.

Geography can also influence the ease of technological adaptation. Technological developments in temperate developed countries, especially in agriculture, are not easily adaptable for use in tropical underdeveloped countries. In *Guns, Germs, and Steel*, Jared Diamond argues that Africa and South America have suffered because their lengths are greater than their widths. Since climate is similar across latitudes, this means that technological developments, particularly in agriculture, were much more able to spread across Europe, Asia, Australia, and North America than South America or Africa. The north-to-south climate and soil variations made it difficult for inventions in Egypt or Argentina to make it to Kenya or Colombia.

Some areas of the world are also more vulnerable to natural disasters. In poor countries without the resources to build structures that can withstand extreme weather events, a single episode can devastate progress achieved over many decades. Increasingly, the differentiated threats posed by global warming in the form of rising sea levels in low-lying nations such as Bangladesh and droughts in the area around Lake Chad will also disrupt growth. They will require expensive diversion of resources simply to mitigate destruction—resources that could otherwise have been used to drive the economy forward.

There is a robust debate among macroeconomists about how important geography is as a fundamental factor relative to institutions. Daron Acemoglu motivates this debate by asking people to consider various pairs of countries or regions, such as North Korea and South Korea. There are vast differences

## 10. What Are Fundamental Barriers to Growth?

in economic and social indicators for the latter country compared with its geographic neighbor. How would economists explain the differences between these countries?

# Institutions

Economists define *institutions* as man-made policies and rules that constrain and shape human interaction. Institutions can be economic (e.g., property rights), political (e.g., methods of electing representatives), legal (e.g., the ability to enforce contracts), or administrative (e.g., corruption of public officials). Investment in physical capital depends critically on having strong property rights. For example, you would not want to start a business if the government could take it away from you by force. Property rights are also vital in providing incentives for a country to use its natural resources wisely. Without property rights, for instance, urban shantytowns with no access to clean water, sewer, or electricity services are common. Strong legal institutions that can impartially settle disputes over property rights or expropriation or business damages are also essential for encouraging domestic investment and attracting foreign investment.

Corruption can seriously dent the ability of entrepreneurs to engage in economic activity and can also result in harmful diversion of investment into projects that do not further the public good. Outside aid can sometimes increase corruption as well. The United States during the Cold War and China in more recent decades have been criticized for using foreign aid and assistance to prop up favored leaders and enhance the donor country's goals. These loans can also create a crippling debt burden that disrupts future growth.

Political institutions matter as well. Even though macroeconomists have shown evidence that greater democracy is associated with faster growth, that is not a settled debate, particularly since the greatest triumph of growth has come from nondemocratic China. However, China had a quite effective political meritocracy for the first few decades after 1978, even within the iron grip of the Chinese Communist Party (CCP). The post-1978 policy of the CCP was to replace leaders every five years and select the rising batch of leaders from the best performers of the next level down. Changes since 2017 by the leader Xi Jinping to abandon the five-year term limits and extend the current leadership's ability to stay in power could disrupt this economic

## 10. What Are Fundamental Barriers to Growth?

success story. Lack of competitive elections and lack of a meritocracy in the administration means that there is more incentive to engage in political machinations rather than acquire education.

Another important institution for growth is a system of intellectual property rights that can encourage innovative activity within the country and also encourage more foreign firms to invest in the domestic economy. The patent extends a temporary monopoly that allows firms to recoup the high fixed costs of developing new ideas and prevents competitors from entering at the low marginal cost stage. However, that does not mean that the best institutions are necessarily always the ones that protect intellectual property rights, especially when it comes to how poor countries deal with foreign inventions. A poor country that does little innovation may choose an IP regime that comes down more heavily on the access side of the access/incentive trade-off than would a rich country that does its own innovation.

Thus, copyright laws often tend to be loosely enforced in less-developed countries because policymakers in those countries see access as the predominant concern. Such behavior can arguably be economically justified. The idea has already been developed, and the patent protection within the rich country has arguably provided the rewards for the inventor. The poor country lacks as clear a trade-off between incentive and access. However, as countries get richer, they may have to change their institutions. For example, it may have made sense for India in the 20th century to focus more on access to technologies rather than incentives to develop technology. India has become a world leader in the manufacturing of generic pharmaceuticals—using chemical formulas developed by others to mass-produce medicine at an accessible price. However, over the years, India has improved its level of technology. It is increasingly at the frontier in areas such as vaccines, pharmaceuticals, and software. Thus, new ideas in those areas will increasingly have to come from India itself. The country must adapt its intellectual property rights regimes to put more focus on incentive over access.

## READING

Collier, Paul. *Breaking the Conflict Trap: Civil War and Development Policy*. World Bank Publications, 2003. <http://documents.worldbank.org/curated/en/908361468779415791/Breaking-the-conflict-trap-civil-war-and-development-policy>.

Diamond, Jared. *Guns, Germs, and Steel: A Short History of Everybody for the Last 13,000 Years*. Random House, 2013.

Easterly, William R. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. MIT Press, 2002.

Robinson, James A., and Daron Acemoglu. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Profile, 2012.

Stiglitz, Joseph E. *Globalization and Its Discontents Revisited: Anti-globalization in the Era of Trump*. W. W. Norton & Company, 2017.

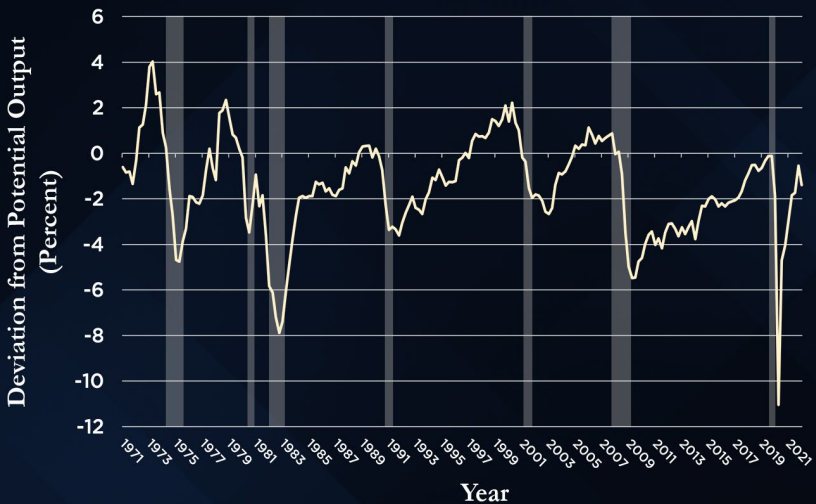
# 11

## HOW RECESSIONS OCCUR AND WHY THEY MATTER

**T**his lecture takes a closer look at economic fluctuations—the sometimes stormy behavior of real GDP around its long-term growth path. In particular, you will be looking at fluctuations of real GDP around its growth trend (i.e., the percentage deviation of real GDP from potential output). Potential output is the output produced by an economy making full, sustainable use of the capital, labor, and technology available to it. Actual output does not always equal potential output. Output can be higher (boom) or lower (downturn) than potential output. The goal here is to understand different perspectives on what causes economic fluctuations.



Fluctuations in the US Economy 1971–2021



## The Keynesian School of Thought

The Keynesian view of economic fluctuations focuses on demand-side explanations for why factors of production are underutilized in recessions. John Maynard Keynes's work sought to understand how such a dramatic downturn as the Great Depression could come about in an economy that had been booming for a decade after the end of the First World War. In the Keynesian view, downturns occur simply because the economy does not make full use of its capital and labor (idling machines, reorganizing workplaces, laying off workers, etc.). Similarly, booms can occur because the economy temporarily makes more intensive use of the available capital, labor, or technology, such as by converting part-time workers into full-time workers.

Keynes illustrated how the interdependency of various sectors within an economy can lead changes in spending to have a multiplied effect on output. A fall in one sector of the economy or region of the country could have a cascading effect on other sectors or regions. This insight also created the

intellectual underpinnings for fiscal and monetary policymakers to play an interventionist role in the economy. If spending decreases by consumers or firms could lead GDP to fall, then the government could step in and increase its own spending to counteract the falling spending. The central bank could step in and make it more attractive for firms and consumers to spend more by lowering interest rates.

A key driving force in this approach is the concept of the marginal propensity to consume (MPC). This parameter describes the fraction of an additional dollar of income that gets spent on additional consumption. Suppose a school hires a new graduate to become a teacher and pays her a salary of \$4,000 per month. The school pays for the services of the teacher, meaning GDP will rise by \$4,000. However, the \$4,000 of expenditure by the school is income to the teacher. For now, assume that the teacher gets to keep the full \$4,000 in additional income. If her MPC is 0.9, she will spend \$3,600 on goods and services, which means that GDP will rise by another \$3,600. That \$3,600 in additional spending by the teacher is additional income to other entities in the economy, such as store owners. If you assume that their MPC is also 0.9, then collectively, their spending will rise as follows:  $0.9 \times \$3,600 = \$3,240$ .

This process will cascade throughout the economy. At each stage, the change in spending equals the MPC times the change in income. The overall effect of hiring a new teacher and paying her \$4,000 is for GDP to rise by much more than the original \$4,000: It's \$4,000 plus \$3,600 plus \$3,240 and so on. The contribution lowers at each stage, but the total is increasing. If you assume that each stage has the same MPC, then the total impact will be \$40,000. This simple calculation is at the heart of the Keynesian approach to economic fluctuations. GDP rises by a multiple of the increase in spending. This magnification is called the Keynesian spending multiplier. In this example, the value of the Keynesian spending multiplier is 10. The same multiplier argument can work in reverse as well—a cutback of \$4,000 in spending would cause GDP to fall by \$40,000.

The intuition behind the model is applicable to the real world. Suppose a large air conditioner manufacturing plant in Indiana shuts down and lays off 1,500 workers. When the company stops paying these workers their salaries, that causes GDP to fall relative to what it had been when they were getting paid. However, the cascading effects will have a much greater negative impact

on the economy. The workers will spend less on eating out and buying new things. Businesses in the town will see their sales fall. They will lay off some of their own workers or cut back on salaries and hours of work. People will move out of town, and house prices will fall. This will reduce property tax revenues, leading the local authorities to hire fewer teachers, police, and firefighters. Over time, the negative economic effects will magnify, and the size of the magnification will depend on the Keynesian spending multiplier.

# Factors Impacting the Keynesian Spending Multiplier

The higher the MPC, the higher the multiplier. In the teacher example, the MPC was assumed to be 0.9. This means that the multiplier was  $1/(1 - 0.9) = 10$ . However, what if the MPC was 0.8? At each stage, the increase in consumption would have been less: \$3,200 instead of \$3,600, for example. In that case, the multiplier would be  $1/(1 - 0.8) = 5$ . In general, the value of the multiplier falls as the MPC falls. If the MPC were only 50 cents of every dollar, then the multiplier would be twice the original amount. If there is less additional spending at each stage of the interconnected links in the economy, spending will rise by less at each stage, and the overall impact will be smaller. The more responsive people's consumption is to income fluctuations, the larger the multiplier will be.

How does the multiplier work in the real economy? For one thing, there are taxes. Consider a 25% marginal tax rate on income. A recipient gets to keep only 75 cents out of every dollar of additional income. With the assumption of an MPC of 0.9, spending will be 0.9 of that 75 cents, meaning 67.5 cents of each additional dollar of income will get spent (as compared to 90 cents in the world without taxes). This is effectively like having an MPC of 0.675. Thus, with income taxes, the spending multiplier would be smaller and equal to  $1/(1 - 0.675)$ , which is 3.07, as compared to  $1/(1 - 0.9)$ , or 10, in a world without income taxes. An MPC of 0.8 will now effectively lead to a multiplier of 2.5 instead of 5, and an MPC of 0.5 will lead to a multiplier of 1.6 instead of 2.

An income tax acts as an automatic stabilizer in the economy. When times are good, with rising incomes, tax payments also rise. When times are bad, with falling incomes, tax payments also fall. The income tax helps to reduce the size

Marginal propensity to consume	Multiplier	Formula
0.90	10	$1 / (1 - 0.9)$
0.85	6.67	$1 / (1 - 0.85)$
0.80	5	$1 / (1 - 0.8)$
0.75	4	$1 / (1 - 0.75)$
0.70	3.33	$1 / (1 - 0.7)$
0.65	2.86	$1 / (1 - 0.65)$
0.60	2.5	$1 / (1 - 0.6)$
0.55	2.22	$1 / (1 - 0.55)$
0.50	2	$1 / (1 - 0.5)$

of the multiplier and keeps economic fluctuations from being as large as they would otherwise be. In deriving the above results, you assumed that the tax rate was 25%. However, in real life, there is a progressive income tax: The marginal income tax rate rises as your income rises. The marginal tax rate for a couple filing jointly that makes \$50,000 is 12%, and their rate at \$100,000 is 22%.

Different types of spending also have different multipliers. In 2015, the Congressional Budget Office did a retrospective study on multipliers from various interventions undertaken in response to the Great Recession of 2009. They looked at nine types of spending, and they provided low and high estimates for each of those multipliers. The first thing to note is that even at the high end, multipliers are nowhere near the value of 10 that you got from the simplest possible formulation of Keynes's idea. At best, high-end estimates are closer to 2. The low-end estimates (and some of the high-end estimates) are less than 1, mostly for transfer payments and tax cuts. A number less than 1 would mean that the increase in GDP is less than the value of the transfer payment or tax cut; that is, more is being saved rather than spent in the economy.

People facing different circumstances are likely to have different propensities to spend. People facing more binding budgetary constraints are more likely to spend additional money. Thus, tax cuts for low-income workers are more likely to have a larger multiplier than tax cuts for higher-income people. Payments to state and local governments to fund ongoing infrastructure projects are more likely to impact spending than allocating a tax credit for first-time home buyers to spend on a new home, since only a small fraction of the people eligible for the tax credit will be tempted to buy a house because of the credit.

The length of fiscal policy changes matters as well. For example, the Trump administration announced an eight-month holiday on collecting payroll taxes between April and December of 2020. In practice, this was likely to have a limited effect for two reasons. First, eight months is a relatively short period of time, meaning it would turn out to be around a 5% cumulative change in take-home pay. Second, since payroll tax cuts are relevant only for people who remain employed, the policy wasn't targeted at people who faced more extreme constraints from income shortfalls. If the announcement had been for a three-year holiday on collecting payroll taxes, the impact would have been greater since that would be a much larger cumulative change in take-home income.

There is also evidence that multipliers are impacted by the state of the economy. When economic conditions are weak and capital and labor are being underutilized, increased government spending or larger tax cuts or transfer payments can encourage firms to hire more workers and produce more goods. In contrast, a similar fiscal intervention when the economy is close to or above potential output will not be that useful since it may only end up causing inflation to rise. Firms that are already using their existing labor and capital quite intensively will find it difficult to expand production in the short run. The Fed's response is also likely to play a role. When there is rising inflation in the economy, the Fed is likely to raise interest rates and try to slow down spending in the economy. This results in the fiscal policy having less of an economic impact.

Overall, the Keynesian approach to economic fluctuations is based on demand-side fluctuations in spending that impact GDP in a magnified way through the spending multiplier. In turn, the proper way to respond to a downturn is also from the demand side. If one sector or area of the economy is cutting back in spending, policymakers should entice some other sector of the

economy to spend more or have the government spend more. The multiplier effect plays a critical role in dictating how much of a response is needed and determining what types of spending increases or tax cuts are needed to boost the economy.

# The Monetarist Theory of Economic Fluctuations

In early 2020, pandemic-related shutdowns of businesses caused spending to plunge and millions of workers to be laid off. However, when the federal government stepped in with expanded unemployment benefits, child tax credit payments, and direct fiscal transfers, consumers began to spend again, albeit on slightly different things (e.g., buying food to eat at home instead of eating at restaurants). What had looked like the return of the Great Depression turned out to be the shortest recession in post-World-War US economic history. If anything, the concern seemed to switch the other way. Did the effort to address high unemployment do too much in terms of increasing spending and providing transfers and tax cuts to people so that the economy ended up in a high inflation situation?

The Keynesian approach rose to prominence during and after the Great Depression, cemented by the more formal work of John Hicks and his model of economic fluctuations, the IS-LM model. Perhaps even more crucial was the work of Paul Samuelson. His highly influential 1948 textbook, *Economics: An Introductory Analysis*, became the dominant reference text for the undergraduate education of a generation of economists. According to the dominant thread in post-World-War macroeconomics, the lesson of the Great Depression was that fluctuations in spending, particularly investment spending, were the key driving force behind economic fluctuations. Policymakers had an active role to play in stabilizing these fluctuations.

In the 1960s, Milton Friedman began to push back on this story. Friedman argued that fluctuations in monetary factors, driven by the mistakes of monetary policymakers, were the driving forces behind economic fluctuations. This claim was based on an influential book he published with Anna Schwartz in 1963. Friedman argued that the pre-Great-Depression approach, in which economic fluctuations were exclusively attributed to

changes in money and credit (especially before the founding of the Fed in 1914), also held true for recessions after the Fed came into existence, including the Great Depression. Friedman showed that monetary aggregates (such as M1 and M2) rose during expansions and fell during recessions.

Friedman and Schwartz closely scrutinized major recessions in the United States up to the mid-20th century and examined the Fed's behavior prior to and during the recessions. They claimed that in 1920, 1931, and 1936/1937, the Fed itself caused recessions by overcorrecting its path. In other words, it would print too much money for a few years and realize that its policies were creating overheating or inflation or inconsistency with the amount of gold reserves it held. Then, it would sharply cut back to such a degree that the economy would be driven into recession. In other cases, such as the onset of the Great Depression, they described the Fed as being too slow to respond with increased money supply because of misconceptions about how much threat was posed to its gold reserves.

This approach led to the emergence of the monetarist theory of economic fluctuations. Monetarists argued for predictable and smooth changes in money supply as the best way to minimize economic fluctuations. The monetarist case was bolstered by Friedman's work on the quantity theory of money. This showed that in a world where velocity growth was predictable and smooth, predictable and smooth growth in the money supply stabilized inflation and minimized real GDP fluctuations. However, subsequent work by monetary economists such as Christina and Paul Romer has shown that using Friedman and Schwartz's approach to selecting supposed Fed mistakes yields other episodes where the Fed mistake did not lead to an economic downturn. Moreover, there are also other, nonmonetary explanations for the recessions that Friedman and Schwartz studied. For example, in 1931, the installation of the Smoot-Hawley tariffs had a contractionary effect on US output and should therefore be responsible for some of the blame apportioned to the Fed by Friedman and Schwartz.

Friedman's work raised enough doubts about the limitations of the Keynesian approach that it led economists to search for alternative explanations for fluctuations. In the 1970s, this led to refinement of both the Keynesian and the monetarist approaches to modeling business cycles. The leading figures were Robert Lucas and Edmund Phelps. Lucas's work, which became known

as the new classical school of economics, emphasized the importance of consumers and firms adapting their behavior in response to policymakers' actions. This meant that a government that tried to follow an overly expansionary fiscal policy, spending too much or taxing too little, would find that its policies became less stimulative over time. Consumers would anticipate that current budget deficits would imply future tax increases and would cut back on current expenditures to prepare for a more financially constrained future. Similarly, monetary policymakers who expanded the money supply to keep spending and GDP higher would end up creating more inflation as firms recognized the implications of this.

# Real Business Cycle Theory

Lucas's view that economic policymakers are limited in what they can do was pushed even further by proponents of real business cycle (RBC) theory. This theory, developed by Finn Kydland and Ed Prescott, grew out of the recession of 1973–1975 and the double-dip recession of the early 1980s. Each recession began with a supply-side shock in the form of a sharp rise in global oil prices. Here, the basic idea is to think of potential output itself as changing over time. In particular, RBC economists typically argue that good and bad shocks to technological progress are the key driving force behind economic fluctuations. Changes such as innovations or an outbreak of war will play an important role in fluctuations, and arguably a larger one than monetary or fiscal policy.

There are two prominent examples of technological changes highlighted by this account that have impacted the US economy in the 2000s. First is the rise of information technology. Second is the subsequent economic transformation from a manufacturing-based to a knowledge-based economy and the increasing importance of intangible capital (e.g., intellectual property rights) over physical capital (e.g., machines).

A different response to Lucas comes from Edmund Phelps. Phelps laid the foundation of the New Keynesian approach to modeling economic fluctuations. This combines the demand-side drivers of economic fluctuations in Keynes's original approach with Lucas-style incorporation of expectations about the actions of policymakers. This led to much more attention to

real-world observations about how firms and workers set prices. These New Keynesian models are widely used in academic circles and in central banks such as the European Central Bank.

## READING

Aghion, Philippe, Roman Frydman, Joseph Stiglitz, and Michael Woodford. “Edmund S. Phelps and Modern Macroeconomics.” *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps* (2003): 3–22.

Hall, Robert E. “Robert Lucas, Recipient of the 1995 Nobel Memorial Prize in Economics.” *The Scandinavian Journal of Economics* 98, no. 1 (1996): 33–48.

Skidelsky, Robert. *John Maynard Keynes, 1883–1946: Economist, Philosopher, Statesman*. Penguin Books, 2013.

The Economist. “What Is the Keynesian Multiplier?” September 8, 2016. <https://www.economist.com/the-economist-explains/2016/09/07/what-is-the-keynesian-multiplier>.

Whalen, Charles, and Felix Reichling. “The Fiscal Multiplier and Economic Policy Analysis in the United States.” Working Paper 2015-02, Congressional Budget Office, 2015. [https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/workingpaper/49925-FiscalMultiplier\\_1.pdf](https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/workingpaper/49925-FiscalMultiplier_1.pdf).

# 12

## AN INTUITIVE MODEL OF ECONOMIC RECESSIONS

**M**odels are like maps. Like a well-designed map, a well-designed economic model can clarify by simplifying. It can help you understand the interrelationships among economic variables and the channels through which a change in one variable can impact others, even if it abstracts considerably from the complexity of the economy. Therefore, this lecture will present an intuitive model of macroeconomic fluctuations.



## Simple Macroeconomic Models

A simple model of how the macroeconomy works can make you more informed. The framework of the model discussed here matches that of models that are used at the IMF and the Fed. Such models are used because macroeconomists cannot run controlled experiments on entire countries.

The history of macroeconomic models is less than a century old. In 1936, Keynes published *The General Theory of Employment, Interest and Money*. The Keynesian view of the economy was formalized by John Hicks's IS-LM model. In this model, there is a spending side, in which spending decisions are impacted by interest rates, and a monetary side, in which changes in money supply by the Fed impact interest rates.

The model used here is a more sophisticated version of the IS-LM model, with a more realistic depiction of central bank behavior than what Hicks assumed. Given that this model is from the Keynesian approach to modeling economic fluctuations, not all economists agree about the conclusions drawn from it. In particular, some economists think that describing aggregate behavior of consumers and firms will lead to misleading conclusions. They prefer an approach where one assumes a prototypical consumer or firm and explicitly describes how that consumer and firm behave. However, this also leads to disagreements about what exactly the merits are of assuming that there exists a representative firm or consumer in the economy.

## Relationships in the Model

This model has two relationships to understand. The first is the aggregate demand curve, which describes a relationship between real GDP and inflation in the economy. The relationship is derived in two steps: first, by determining how changes in the real interest rate affect real GDP, and second, by determining how the real interest rate itself is determined by the actions of the Fed in responding to inflation. The real interest rate is the nominal interest rate minus inflation. It is the best measure of the true cost of borrowing (or the true benefit from saving). Suppose you took out a consumer loan that has an interest rate of 18%. In the United States, for every \$1,000 you borrowed, you would have to pay \$1,180 back after a year. However, suppose you lived in

## 12. An Intuitive Model of Economic Recessions

Turkey, where inflation in 2022 shot up to 45% and kept rising. Each unit of currency you repaid after a year would have only a little more than half of the purchasing power of the money you borrowed. You would be getting a good deal if you could get a loan at 18% interest.

By correcting the nominal interest (18%) for inflation (45%), the real interest rate in Turkey would be  $-27\%$ , meaning the true cost of borrowing is negative. In high-inflation Turkey, an 18% interest rate would be great for the borrower and terrible for the lender. In contrast, in the United States, where inflation has historically been around 2%, the real interest rate would be 16%—a great deal for the lender but less so for the borrower. Therefore, what truly matters for spending decisions in the macroeconomy is the real interest rate. There is a negative relationship between the real interest rate and real GDP. Higher real interest rates reduce spending by making it expensive to borrow, and the opportunity cost of spending your own money increases (because saving the money instead of spending it will be more attractive). Lower spending leads to lower GDP. Conversely, lower interest rates lead to more spending and higher real GDP by making it cheaper to borrow and reducing the incentive to save.

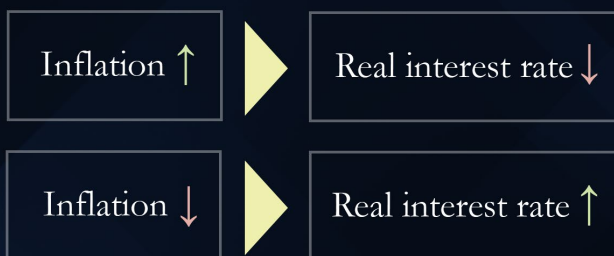
If the nominal interest rate is **18%**  
and inflation is **45%**,  
the **real** interest rate is  **$18\% - 45\% = -27\%$**

If the nominal interest rate is **18%**  
and inflation is **2%**,  
the **real** interest rate is  **$18\% - 2\% = 16\%$**

Consider how the Fed manipulates the real interest rate in response to inflation. Based on work done by John Taylor in the early 1990s, among others, most economists now feel comfortable in generally describing the Fed's

behavior using a simple policy rule. An important component of this policy rule is describing how the Fed changes the real interest rate when inflation deviates from its target rate of inflation. Typically, the Fed has used 2% as its stated target for inflation.

When the Fed thinks inflation is too high relative to its target, it attributes this to the fact that there's "too much money chasing goods." It raises the real interest rate in the hope that this will lead to less spending and that less money chasing goods will drive inflation down. Conversely, when the Fed thinks that inflation is too low, it attributes this to the fact that there's "too little money chasing goods." It then lowers the real interest rate in the hope that this will lead to more spending and that more money chasing goods will drive inflation up.



## Deriving the Aggregate Demand Curve and the Inflation Adjustment Line

The aggregate demand (AD) curve is a relationship between inflation and real GDP. Putting the above relationships together, you get an inverse relationship between inflation and real GDP. Consider an economy that has an inflation level  $\pi_0$  and a GDP level  $Y_0$ . Suppose inflation decreases to a value  $\pi_1$  that is

## 12. An Intuitive Model of Economic Recessions

less than  $\pi_0$ . A fall in inflation will lead the Fed to lower the real interest rate. This will lead to more spending, which will raise real GDP. Thus, real GDP will increase to  $Y_1$ , which is greater than  $Y_0$ .

Conversely, suppose inflation increases from  $\pi_0$  to  $\pi_2$ . This will lead the Fed to raise the real interest rate. Moreover, it will lead to less spending, and less spending will lower real GDP. Thus, real GDP will decrease to  $Y_2$ , which is less than  $Y_0$ . You end up with a downward-sloping AD curve. This shows the relationship of inflation movements to fluctuations in real GDP. What causes this curve to shift? The key intuitive idea is that a rightward shift of the AD curve represents changes that cause spending and real GDP to rise. Conversely, an inward shift represents changes that cause spending and real GDP to fall.

The second part of this model of economic fluctuations is the inflation adjustment (IA) line. This is a horizontal line depicting the current inflation rate in the economy. Three key assumptions are made about how inflation adjusts. The first assumption is that unless there are unexpected supply shocks, the inflation rate does not move much in the short run. This assumption is motivated by the fact that any individual wage or price in the economy is adjusted infrequently compared to the moving total of all price changes in the economy. Wages adjust infrequently because workers in many industries have contracts that set their wage for a specific period of time. Moreover, firms tend to adjust prices infrequently. This is partly because their wage costs are under control, partly because their other costs are also governed by contracts to buy goods (e.g., raw materials), partly because their prices have already been set in contracts to supply goods to retailers, and partly because of menu costs (i.e., the physical costs of changing prices). Thus, typically, in the absence of supply shocks, even if real GDP were to change suddenly, it would take some time for inflation to respond.

The second assumption is that over time, inflation will rise or fall depending on where real GDP is relative to potential GDP. Over a longer period of time (e.g., over a year or so), inflation will change. When the economy is doing better than typical, people are spending more on goods and services, incomes are higher, and unemployment is lower. The tight labor market drives up wages. Workers can quit and find another job more easily, putting pressure on firms to pay higher wages. Firms can handle these higher costs because there is robust demand for their goods and consumers are more willing to

pay higher prices for goods. Thus, wage and price contracts are likely to build in larger increases. As a result, inflation will rise. Conversely, if the economy is doing worse than typical, people are not spending on goods and services, incomes are lower, and unemployment is higher. Workers lose bargaining power and accept lower wages because it may be difficult to find a new job. Firms have to keep price increases low because consumers are being more cautious in their spending habits. Thus, wage and price contracts are likely to build in smaller increases. As a result, inflation will fall.

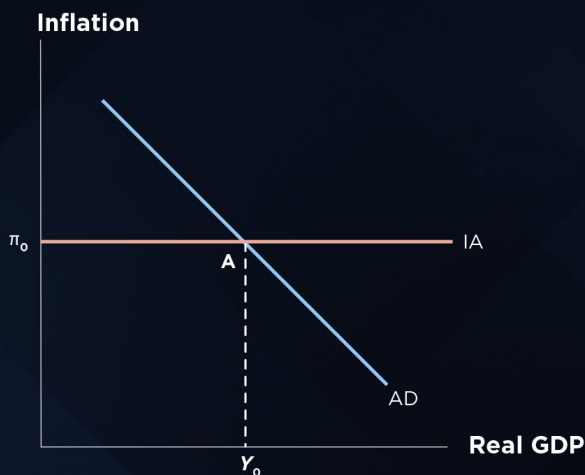
The third assumption is that supply shocks can cause inflation to rise or fall sharply, even in the short run. The typical supply shock would be a sharp rise or fall in the price of oil. A sudden jump in the price of oil will have an economy-wide ripple effect in raising prices: Prices of almost all goods are affected in some way by the price of oil, either in production or in transportation. Therefore, a sudden jump in the price of oil can cause prices to rise more sharply and thus inflation to be higher than it has been in the recent past. However, the pandemic has provided another example of a supply shock: a public health crisis that throws complex supply chains out of whack. When a firm faces a sharp rise in input costs, it will try to pass those costs on to its customers by raising the price of its goods or renegotiating some of the existing contracts in the economy.

## Combining the AD Curve and IA Line

Recall that the AD curve shows how real GDP responds to different inflation rates. The IA line shows how inflation evolves in the economy. Thus, the intersection of the two gives the current rate of inflation ( $\pi_0$ ) and the current level of real GDP in the economy ( $Y_0$ ). The intersection can appear either higher or lower or even exactly at the level of potential output in the economy. Since inflation can change over time depending on where real GDP is relative to potential output, that intersection is only a short-run equilibrium. Over time, inflation will rise or fall. Potential output is indicated by  $Y^*$ .

Consider three cases. In case 1, where the economy is doing better than typical (GDP exceeds potential output), inflation will rise over time, and the IA line will shift upward. The intersection of AD and the rising IA show that

## 12. An Intuitive Model of Economic Recessions



real GDP will fall over time until it reaches potential output. Conversely, in case 2, where the economy is doing worse than normal (GDP is less than potential output), inflation will fall over time, and the IA line will shift downward. The intersection of AD and the falling IA show that real GDP will rise over time until it reaches potential output. Finally, in case 3 (where GDP equals potential output), the IA line will not shift over time, and real GDP will remain at potential output.

The key feature of the model is the self-correcting nature of the economy. When spending gets too high relative to what the economy can sustainably produce (GDP is higher than potential output), inflation rises. This leads the Fed to raise real interest rates, reduce spending, and bring real GDP back down toward potential output. Meanwhile, if spending falls and the economy goes into a downturn (real GDP is less than what the economy can sustainably produce), inflation falls. This leads the Fed to lower real interest rates, increase spending, and bring real GDP back up toward potential output. Only when real GDP ends up equaling potential output will inflation not rise or fall any further. This will be the long-run equilibrium of the economy.

## Common Macroeconomic Scenarios

In this analysis, assume that the change being analyzed is the only change in the economy. This assumption of holding all else equal is the *ceteris paribus* assumption. In addition, assume that the economy is starting out in long-run equilibrium with real GDP equal to potential output.

Consider a tax cut, which will lead to higher spending. The AD curve will shift right. The tax cut initially raises real GDP but does not immediately raise inflation. However, since the short-run equilibrium is above potential output, inflation will rise over time. As inflation rises, the IA curve will shift up, and real GDP will fall. (Intuitively, you are moving back along the AD curve because the Fed is raising interest rates in response to the rising inflation.) This will continue until the economy has returned to potential output.

Next, consider the impact of a stock market crash that wipes out the value of consumers' retirement portfolios and causes a sharp fall in consumer confidence. Consumers will reduce their spending, and the AD curve will shift inward. Since the short-run equilibrium is below potential output, inflation will fall over time. As inflation falls, the IA curve will shift down, and real GDP will rise. (Intuitively, you are moving forward along the AD curve because the Fed is lowering interest rates in response to the falling inflation.) This will continue until the economy has returned to potential output.

Now, consider disruption in supply chains as a result of COVID-19. This is represented as an upward shift in the IA line in an economy that is at  $Y^*$ . The supply shock immediately causes inflation to rise. Real GDP is lower in the short-run equilibrium because the higher inflation rate has led the Fed to raise the real interest rate, which in turn has brought real GDP down. Over time, since the short-run equilibrium real GDP is lower than potential output, inflation will gradually fall. As inflation falls, the IA line will shift down, and real GDP will rise. (Intuitively, you are moving back along the AD curve because the Fed is lowering interest rates in response to the falling inflation.) This will continue until the economy has returned to potential output.

## 12. An Intuitive Model of Economic Recessions

This also illustrates why supply shock inflation is so economically damaging. In this case, the economy is experiencing both lower real GDP (a stagnant economy) and a much greater increase in prices (higher inflation). This combination is called stagflation and was a constant economic threat for the United States in the 1970s and 1980s. Also note that although, in the long-run equilibrium, real GDP is equal to potential output and potential output is a constant vertical line labeled  $Y^*$  in the model, in the real world,  $Y^*$  is changing over time; that is, potential output is the long-run growth path of the economy. Saying that real GDP equals potential output in the long run is not a statement that real GDP is constant in the long run. The model is saying that real GDP ends up on its growth path in the long run, although, in the short run, various forces can push it higher or lower than its long-run growth path.

All examples up to this point assume that the economy started out at potential output. If that is the case, outward shifts of the AD curve seem to lead to rising inflation. Inward shifts seem to lead to falling inflation. However, recall that IA reflects whether real GDP exceeds, equals, or falls short of potential output—whether it is at, below, or above  $Y^*$ . For example, suppose the economy starts out in a downturn. The Fed wants to intervene by lowering the path of interest rates to boost real GDP in the economy. In other words, it pushes the AD curve to the right. Suppose that the magnitude of the outward shift is such that the short-run equilibrium of the economy is still lower than potential output. Then, even though the Fed has pushed the AD curve to the right, over time, inflation falls rather than rises. The inflation rate would have fallen even more, given the starting position, if the AD shift had not happened.

Another possibility is that the Fed and the government did a little too much and ended up pushing AD beyond potential output. This would lead to the economy settling at a long-run equilibrium with higher inflation. This is the policy debate that occurred in late 2021 as the economy recovered from the COVID-19-induced recession. Something needed to be done to stave off another decade of economic underperformance as the economy tried to recover from that shock. Both fiscal and monetary policymakers stepped in and tried to push AD higher. The debate was about whether they did too little or too much.

## READING

Krugman, Paul. “Mr. Keynes and the Moderns.” Prepared for the Cambridge conference commemorating the 75th anniversary of the publication of *The General Theory of Employment, Interest, and Money*. June 18, 2011.

Romer, David H. “Keynesian Macroeconomics without the LM Curve.” *Journal of Economic Perspectives* 14, no. 2 (2000): 149–169.

Stewart, Emily. “The Risks of Going Too Big on Stimulus Are Real.” *Vox*, February 8, 2021. <https://www.vox.com/policy-and-politics/22268787/larry-summers-op-ed-biden-stimulus>.

The Economist. “A Working Model: Is the World Experiencing Excess Saving or Excess Liquidity?” August 11, 2005. <https://www.economist.com/finance-and-economics/2005/08/11/a-working-model>.

Weerapana, Akila. “Intermediate Macroeconomics without the IS-LM Model.” *The Journal of Economic Education* 34, no. 3 (2003): 241–262.

# 13

## LESSONS FROM PAST RECESSIONS

**M**any of the major concepts used for analyzing and responding to economic recessions come from macroeconomists' understanding of the Great Depression. For contrast, this lecture will also look at the 2001 recession, which was the mildest postwar recession the United States has experienced. The policy response to that recession well reflected the knowledge gained from the policy mistakes of the Great Depression. Then came the 2008/2009 downturn, or the Great Recession. Understanding why the 2008/2009 recession was so severe, and why the policy response was unable to alleviate the depth and severity to the extent it could in 2001, sheds light on the 2020 COVID-19 recession and the even larger and more aggressive policy response to it.

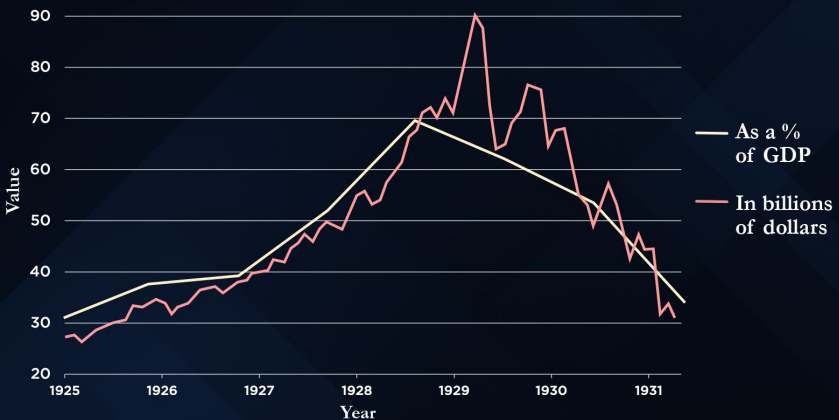


# The Great Depression

The Great Depression of 1929 to 1933 was the most severe economic downturn that the United States faced. To best understand what caused it, consider the period before 1929 and the decade known as the Roaring Twenties. The Roaring Twenties saw a rapid growth in the economy and a booming stock market. Initially, the boom in stocks was driven by the end of World War I, the rapid rise in the electrification of cities, and new inventions such as the motor car and the telephone. All seemed to herald a new era of economic prosperity. Data from Angus Maddison illustrates the Roaring Twenties (and the subsequent crash) vividly: Real GDP rose from \$750 billion in 1921 to more than \$1.1 trillion at the time of the crash.

The initial euphoria over a booming economy and a rising stock market turned into overoptimism and speculation as more people started to buy stocks in the belief that the prices would continue to rise. However, ever-rising stock valuations become harder to justify over time. These episodes are called bubbles, and all bubbles eventually pop. At some point, confidence fails, and what had been a rush to enter the stock market quickly becomes a panicked

Market Value of Companies on the NYSE, 1925–1931



rush to the exit. In October 1929, the stock market crashed, beginning on Black Tuesday, and fell by more than 60% over the next month. This wiped out close to \$30 billion of net worth (30% of pre-crash GNP) and destroyed consumer and investor confidence.

In 1929, the exposure to a stock market bubble pop was compounded by high degrees of leverage: Banks and brokers had lent \$9 for every \$1 that the customer would put up, enabling people to buy stocks in amounts that far exceeded their own liquid resources. Suppose you had \$1,000 to invest in the stock market. The market produces a 20% gain in one year, meaning you will have \$1,200 now invested in the market. You have gained \$200, a return of 20% on your original investment. Suppose that the next year, there is a 20% loss in the market. The value of your portfolio falls by  $\$1,200 \times 20\% = \$240$ , leaving you with \$960. You have now lost 4% of your initial investment.

Now, consider the same scenario under leverage. Suppose that a bank would lend you \$9,000 for every \$1,000 you had. Then, you could invest \$10,000 into the market. The first year, when the market rose by 20%, your portfolio would rise to \$12,000. You have earned \$2,000 on your \$1,000 investment—a 200% increase rather than a 20% increase. If you cashed out in time, you could pay back the bank and have three times as much money as you started out with. However, if you stay in the market and experience a 20% loss, the value of your portfolio falls by  $\$12,000 \times 20\% = \$2,400$ , leaving you with \$9,600. You have now lost 40% of your initial investment. David Wheelock stated that the crash was 80%; in this case, you would have lost \$9,600 of your portfolio, leaving you with \$2,400. Remember that you owe \$9,000 to the bank, meaning that not only have you lost all \$1,000 of your own money but you also owe the bank \$6,600.

Banks that had made leveraged loans to speculators suffered as speculators defaulted on these loans when their stock market investments cratered. The result was a sharp cutback in bank lending that further eroded investor confidence. To add to the panic, there were bank runs. A bank run begins when depositors panic that the banks may be unable to return their money. This results in a rush of people withdrawing their money from the banks, creating liquidity problems. This causes a sharp rise in interest rates as banks try desperately to hold on to deposits.

The policy response to the Great Depression only made things worse. Brad DeLong described fiscal policy in that era as the federal government trying to balance wartime deficits with peacetime surpluses. This differs from the more conventional countercyclical fiscal policy that the government uses today, where it pursues more expansionary policies in bad times and more contractionary ones in good times. At the outset of the Great Depression, coming off a widespread belief that the downturn was a necessary “cleansing” of the excesses of the Roaring Twenties, the appropriate fiscal policy response at the time was understood as being to “keep the budget balanced and live within the means of the government.” This meant raising taxes and cutting government expenditures, which would further stifle aggregate demand.

There is vigorous debate about whether monetary policy in this period was neutral or contractionary. However, there is no doubt that it was not expansionary. At this time, the United States was on the gold standard. This meant that the Fed could not expand the money supply without acquiring more gold reserves. Christina Romer and Barry Eichengreen have argued that the Fed was seriously constrained by the gold standard after 1931. Great Britain’s decision to exit the gold standard had already caused a fall in US gold reserves because of the expectation that the United States would soon follow suit or at least devalue the value of the dollar against gold.

A third key policy move during this period was the decision to embark on the Smoot-Hawley tariffs in the misguided belief that shutting down imports would stimulate demand in the domestic economy. Instead, other countries retaliated, further reducing GDP in the economy. Data from the BLS shows deflation of between -5% and -10% persisting for more than two years during the Great Depression.

## The End of the Great Depression

The Great Depression ended when the extraordinarily contractionary policies were changed. First, the United States came off the gold standard, which allowed for a more expansionary monetary policy. The election of President Roosevelt brought his support for expansionary fiscal policy via the New Deal programs. In 1934, the Smoot-Hawley tariffs were disbanded. A rapid rise in government military expenditures as the economy geared up for entry into World War II also helped bring the recession to an end in the late 1930s.

John Maynard Keynes was a resounding critic of the early contractionary monetary and fiscal policy responses to the Great Depression. His writings were formalized by John Hicks, resulting in the development of the Keynesian model and the IS-LM model, which became the cornerstone of macroeconomics in the 1950s and 1960s. Those decades also saw the ascendancy of Keynesian economists who used these models to advocate for a much more activist role for policymakers, especially in downturns. Those models were predecessors of the AD-IA model.

Moreover, the Glass-Steagall Act forbade commercial (deposit-taking) banks from investing in stock markets. The Federal Deposit Insurance Corporation (FDIC) insured consumer deposits at commercial banks up to \$2,500 in 1933 or up to \$250,000 now. Also important was the formation of the Securities and Exchange Commission, which passed many new rules to “maintain fair, orderly, and efficient markets.”

## The 2001 Recession

The United States experienced a dozen recessions from 1945 through the COVID-19 recession of 2020. None came anywhere close to the magnitude of the Great Depression. The progress seemed to be particularly noticeable in the 15 years between 1992 and 2007, during which the economy suffered only one recession in 2001. Lasting from March to December 2001, this was the mildest postwar recession when measured in terms of the decline in GDP.

However, there are some interesting parallels between this recession and the Great Depression. The 1990s, particularly the latter half, was a decade of rising economic prosperity in the United States. The rise to prominence of the IT sector and the widespread use of the internet offered the promise of a fundamentally more prosperous future, just as the Roaring Twenties had promised. Moreover, like in the Roaring Twenties, there was a rapidly booming stock market, particularly the technology-heavy NASDAQ index, which increased by 400% between November 1995 and November 1999. New internet companies were able to attract millions of dollars in venture capital financing. Even existing companies expanded or created web-based parts of their business. Meanwhile, investors who were able to buy shares of these nascent technology companies found themselves fabulously wealthy as stock valuations boomed.

By around 2000, it became clear that many of these companies could never justify the crazy valuations that had been placed on them. Furthermore, many of the founders and early investors who held shares were often restricted from selling shares until a certain time period had passed. As these restrictions expired, these early investors started cashing out. Soon, the rush to enter turned into a rush to exit. As in the Great Depression, the initial impetus was a stock market crash that drove down investor and consumer confidence.

The similarities end there. Unlike in the Great Depression, there were no widespread bank failures. Although the Glass-Steagall Act restrictions had been removed in the late 1990s, banks had not moved into speculative stock purchases in the way that the 1920s banks had. Furthermore, the presence of deposit insurance ensured that there was no run on the banking sector and no panic among depositors.

The policy response to the 2001 recession was also strikingly different. The Fed, unburdened by the gold standard, aggressively expanded the money supply and drove interest rates lower. The new president, George W. Bush, had advocated during his campaign for a tax cut returning the late 1990s budget surplus to taxpayers. Even though the surplus dwindled when the economy went into recession, he did not change his policy. He instead argued for the tax cut as a needed boost for the economy while also increasing military and security spending after 9/11. The combination of expansionary fiscal and monetary policy fundamentally differed from the initial response to the Great Depression, and the economy recovered fairly quickly.

## The Great Recession

However, in 2007/2008, the Great Recession hit. The precipitating event was initially a collapse in housing prices in a few states where the market had risen rapidly: Arizona, Florida, California, and Nevada. This was followed by a stock market collapse headlined by the failure of two investment banks, Bear Stearns and Lehman Brothers. They had invested extensively in assets that derived their value from the underlying prices in the real estate markets.

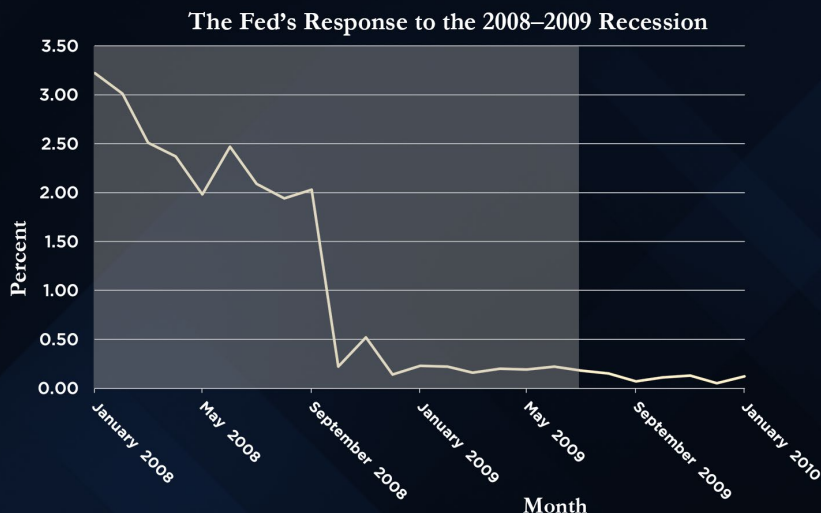
Once again, there had been an asset price collapse that impacted consumer spending. Unlike in 2001, though, there was a collapse in the overleveraged banking sector. Moreover, unlike in the Great Depression, the collapse was not

because of overinvestment in the stock market but because of overexposure—with too much leverage—to the declining real estate market. There was an even more dramatic decline in the non-bank financial sector that had invested in exotic financial instruments related to housing markets. This resulted in a sharp fall in investor confidence as access to credit became limited.

This time, however, there was no broad-based panic because deposit insurance protected individuals. Deposit insurance was expanded to \$250,000, and depositors were able to stay confident that deposits were secure. The importance was further underlined by the panic in the non-bank financial sector. People who had money in hedge and money market funds, which were not insured, panicked and started to withdraw money. Thus, overall, the magnitude of the 2008/2009 shock was more severe than in 2001 but less so than in the Great Depression.

The federal government reacted with a substantial expansionary fiscal policy in the form of a \$400 billion tax cut presented by the outgoing Bush administration. This was followed by an \$800 billion stimulus program designed by the Obama administration. The expansionary fiscal policy at the federal level was muted by the contractionary fiscal policy at the state and local levels. Ironically, many of those governments operated under balanced budget rules that made their behavior resemble the policies enacted early in the Great Depression, pursuing contractionary rather than expansionary policies in a period of economic contraction. However, even the Fed's expansionary monetary policy was limited in its effectiveness by hitting the zero lower bound on nominal interest rates. Since such policy boosts the economy by increasing investment via lower interest rates, limits on how much interest rates can fall can disrupt the monetary policymaker's ability to boost the economy.

The general uncertainty level did not help the economy either. Tax cuts require that people spend them on goods and services for generating economic activity. More money supply requires that banks be more willing to make loans to increase economic activity. In 2008/2009, people were more interested in saving rather than spending money. Banks were more interested in holding on to reserves rather than making loans, which reduced the effectiveness of these policies. The end result was a downturn that was much more severe than the post-World War II recessions.



After 2009, most of the concern focused on three areas of policy responses: how to build a more resilient banking sector, how to conduct monetary policy if the lower bound of zero was reached on interest rates, and how to coordinate an effective fiscal policy response if monetary policy was constrained.

## COVID-19

In March 2020, signs of rapid increases in COVID-19 led to governors and mayors implementing emergency measures that closed schools and nonessential businesses and restricted travel. This led to a dramatic shutdown in economic activity. Consumer spending fell sharply. Many businesses thus had to furlough or lay off workers. Investment also fell sharply, with sharp decreases in purchases of new equipment and buildings to facilitate production. Bankruptcies at large firms, such as Hertz, appeared frequently in the news.

The COVID-19 pandemic was a global disaster. Thus, international trade was severely disrupted, and US exports fell. In the early days of uncertainty, stock markets also reacted negatively. Financial markets declined sharply in

March 2020, substantially eroding the wealth and retirement portfolios of consumers. There was a substantial rise in uncertainty, even more so than in the Great Recession.

Using the model, you would predict the combined effect of these changes to be a sharp fall in real GDP and in inflation. The data matches these predictions. On the output front, the cumulative impact was the most damaging shock since the Great Depression, albeit limited to the second quarter of 2020. There was also a substantial fall in inflation driven by lower energy prices: As global economic activity declined and air travel virtually shut down, demand for gasoline and crude oil plummeted. Simultaneously, a price war broke out between major oil-producing nations, resulting in an increase in global supply. This sent crude oil prices falling by more than 60% in the early months of 2020. This resulted in the price of a gallon of gas falling by 25% between January and April.

Fortunately, the 2020 recession also turned out to be the shortest modern recession. First, the stock market not only recovered its losses but also boomed in the latter part of 2020 and into 2021 as it became clear that the pandemic was creating both winners and losers. The winners included some of the largest companies in America, such as Amazon, Apple, and Google. Second, vaccines were developed and rolled out to the population. These vaccines reduced uncertainty—at least in the short term—about the health costs of the pandemic and the negative impacts on economic interactions. Much of the recovery came from fiscal and monetary policy responses that built on the experiences of the three aforementioned recessions.

As in 2008, the Fed acted quickly in 2020 to lower interest rates as low as they could go. The Fed typically makes decisions to change interest rates at scheduled meetings of the FOMC. In March 2020, the FOMC called two unscheduled meetings in a single month. The first meeting cut the federal funds rate by 50 basis points (half a percentage point) in early March. The second meeting lowered rates by another full percentage point, until the lower bound of zero was reached. On the fiscal policy side, Congress passed the CARES Act. The \$2 trillion package supported payments to families, loans to companies, assistance to hospitals, and the expansion of unemployment insurance payments. All of this helped consumers and firms meet essential expenditures while the economy was shut down.

The economy recovered rapidly from the initial cataclysmic shock of the pandemic. You would think that there would be broad-based agreement that the policy response to the 2020 recession was appropriate. However, there was disagreement. In particular, the argument was that by doing too much, policymakers created more inflation in the economy—especially when the original policy responses were augmented by more expansionary fiscal policy from the outgoing Trump administration in late 2020 and by the incoming Biden administration.

## READING

Bernanke, Ben S. *Essays on the Great Depression*. Princeton University Press, 2009.

———. *The Federal Reserve and the Financial Crisis*. Princeton University Press, 2013.

Blinder, Alan S. *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead*. Penguin Books, 2013.

Lowenstein, Roger. *Origins of the Crash: The Great Bubble and Its Undoing*. Penguin, 2004.

Williams, Heidi, ed. “Symposium: COVID-19.” *Journal of Economic Perspectives* 35, no. 3 (Summer 2021): 3–66. <https://www.aeaweb.org/issues/646>.

# 14

## WHAT IS GOOD MONETARY OR FISCAL POLICY?

**W**hat constitutes good macroeconomic policy? Despite the fact that disagreements about policy play out regularly in the media, there are areas in which macroeconomists can come to agreement based on sound economic judgment. This lecture will use the macroeconomic model to guide your thinking while acknowledging the limits of such a model-based approach. It will also illuminate why real-world policymaking can be so challenging and contentious. You will be analyzing both fiscal policy (tax changes and government spending changes) and monetary policy (interest rate changes by the Fed).



## What Is Fiscal Policy?

Recall that expansionary policy aims to boost spending in the economy, which involves tax cuts and/or spending increases on the fiscal side and a lower path for interest rates on the monetary side. Neutral policy is the status quo: The Fed follows the current interest rate path, and the government sticks to existing tax and spending paths. Finally, contractionary policy aims to reduce spending in the economy. This involves tax hikes and/or spending decreases on the fiscal side and a higher path for interest rates on the monetary side.

	<b>Expansionary = Boost Spending</b>	<b>Neutral = Status Quo</b>	<b>Contractionary = Reduce Spending</b>
<b>Fiscal Policy</b>	Tax cuts and/or spending increases	Government sticks to existing tax and spending paths	Tax hikes and/or spending decreases
<b>Monetary Policy</b>	A lower path for interest rates	Fed follows the current interest rate path	A higher path for interest rates

Consider economic policy in a downturn. A downturn is broadly defined as a recession, which is a period when output is broadly declining or GDP is below potential output. The actual recession is a period of falling real GDP. Although the 2008/2009 recession was officially deemed to have ended at the end of 2009 when GDP began to recover, it was not until the first quarter of 2018 that real GDP caught up to potential output again. Thus, the definition of a downturn here includes the recovery period from a recession, when the economy has not yet returned to its long-run growth path.

During a downturn, a contractionary policy, whether it be fiscal or monetary, is not a good idea. It would move the economy even deeper into recession and lead to a longer recovery and a greater threat of inflation being too low and potentially falling into deflation. Despite this, the clamor for fiscal austerity and belt-tightening in a downturn has been seen in the United States, the United Kingdom, Greece, and Ireland in the last decade.

#### 14. What Is Good Monetary or Fiscal Policy?

The second possibility for a policy response to a downturn is a neutral policy. Over time, inflation will start to fall. Eventually, the economy will return to potential output with a lower inflation rate. This is a better outcome than the contractionary policy. However, every period that the economy is passively allowed to stay in recession means that people are without jobs and the economy is underperforming.

Thus, based on the simple macroeconomic model, the best policy seems to be for the government and/or the Fed to act decisively to pursue the third possibility, an expansionary policy, to move GDP back to potential output as quickly as possible. However, is it better to use fiscal policy or monetary policy? If fiscal policy is used, should it take the form of increased government purchases or tax cuts or a combination of both? If government purchases are increased, what should be increased? If taxes are to be cut, whose taxes should be cut? A simple model cannot answer all of these questions.

## The “Three Ts”

Larry Summers’s characterization of good policy is to always incorporate the “three Ts”: “targeted, timely, and temporary.” Targeted policy has a large multiplier effect so that GDP can be changed with as little a cost as possible. For example, in a recession, to raise GDP, spending should be increased or taxes cut in ways that have as large an economic impact as possible. In 2020, the Trump administration sent out \$2,400 checks to households under the CARES Act. The administration decided to restrict the payment to households earning less than \$150,000. Why? Mostly because financially constrained households were more likely to spend the CARES act transfers rather than save them.

Targeting monetary policy appropriately is even harder because the interest rates that any one individual or firm faces can’t be lowered. When the Fed resorted to extraordinary policy measures after hitting the zero lower bound in the 2008/2009 and 2020 recessions, it was criticized for using policies that were designed to stabilize and boost financial markets. This would implicitly safeguard the retirement and investment portfolios of the well-off rather than help, for example, a small restaurant struggling to access credit to keep afloat.

#### 14. What Is Good Monetary or Fiscal Policy?

Timely policy is policy that kicks in as soon as possible. This way, people do not have to suffer the impact of the recession or downturn more than necessary. This means that a new pipeline project in South Dakota that requires many rounds of studies and lawsuits before being built is a much worse fiscal policy to pursue in a recession than assistance to rebuild a failing bridge in a major city.

Many economists argue that expansionary policy interventions that are designed to nudge the economy back onto its long-run growth path should be left in the hands of the Fed rather than Congress. Getting a policy through Congress can take many months or even years. If the president vetoes the bill, the whole process has to start over again. In addition, at every stage of the legislative process, amendments make changes to the original bill. Thus, by the time the legislation is finally enacted, the economic circumstances may have changed so much that the bill is no longer the appropriate response. Furthermore, reversing a fiscal policy once enacted can be just as challenging. If future circumstances call for a contractionary fiscal policy, undoing the current expansionary policy may take just as long as putting it into place initially did.

In contrast, the Fed requires no congressional approval and can move quickly and decisively based only upon a vote of its FOMC. Further, it can take away what it gives to the economy just as easily: Rate hikes can be implemented as easily as rate cuts can. What the Fed faces are impact lags—time for the policy to impact the economy. Consumer and investor responses to rate changes are likely to take a few months or more. Thus, it may take time for the impact of a Fed policy decision to be felt in the economy and even longer for the impact to be fully realized.

Recent research by David Gruen, John Romalis, and Naveen Chandra has estimated an average lag of five to six quarters for monetary policy to have a full effect. Thus, central banks that have an explicit inflation target are only responsible for achieving that target over a relatively long horizon (e.g., two years). Overall, though, economists broadly agree that discretionary monetary policy is more nimble than discretionary fiscal policy. As fiscal policy is less nimble, economists agree that it works best in the form of automatic stabilizers: built-in shock absorbers that can limit the size of economic fluctuations in both directions, such as unemployment benefits. Such programs provide support to households when their incomes fall and

#### 14. What Is Good Monetary or Fiscal Policy?

withdraw their support when incomes rise. Since this happens automatically, it works more effectively than discretionary policy that is initiated in response to a particular economic situation.

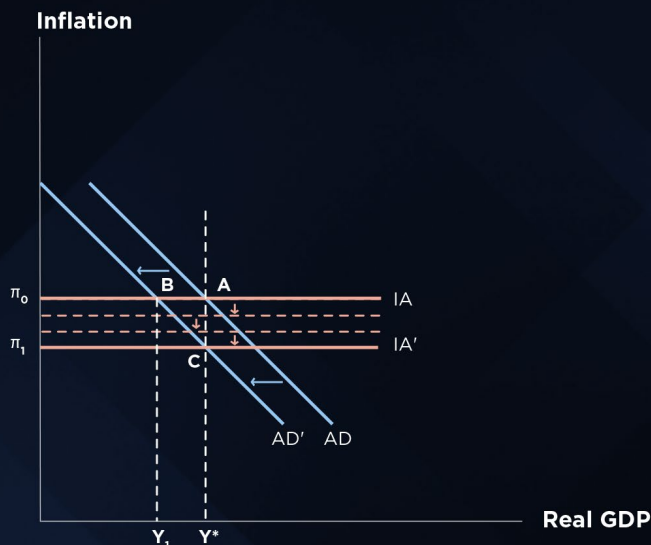
The third T stands for *temporary*. The key idea here is to avoid implementing policies whose effects will linger into the future, when the context changes so much that the policy is inappropriate. As an example, to increase government expenditures during the COVID-19 recession, a good option would be a one-time expenditure on a COVID-19 vaccine affordability program. On the tax side, putting in a new tax bracket at the low or high end of the existing tax system is likely to be a much more long-lasting intervention than the one-time \$2,000 child tax credit that the Biden administration passed in the American Rescue Plan in 2021.

## Short-Term Booms

What happens when real GDP exceeds potential output—when the economy is in a short-term boom? The notion that an economy can be too strong is often a difficult one for people to understand. A short-term boom has long-term negative consequences in the form of higher inflation. If you follow expansionary policy, which is the policy of choice when the economy is in a downturn, this will result in a worse outcome in the long run. The economy will end up back at potential output, with inflation having climbed much higher.

A second possibility is to follow a neutral policy; over time, inflation will start to rise. Eventually, the economy will return to potential output with a higher inflation rate. Although this is better than the expansionary policy, once the boom has dissipated, there is higher inflation. Thus, the benefits are temporary, while the costs are permanent.

The third possibility is to follow a contractionary policy that moves the economy back to potential output. This can take the form of higher taxes and/or lower spending from fiscal policy or a higher path for interest rates from monetary policy. This is the most suitable policy when the economy is in a boom.



## Real GDP Equaling Potential Output

Consider the possible policy options when real GDP is equal to potential output. What if you follow a contractionary policy, which is best for the economy when in a boom? This will drive the economy into a recession in the short run and lead to a fall in inflation over time. What if you follow an expansionary policy? That's the recommended policy choice for an economy when it is in a downturn. This will create a boom in the short run, which seems desirable. However, since you are pushing spending higher without boosting the economy's productive capacity, the policy is unsuccessful in the long run since it causes higher inflation.

The third possibility is to follow a neutral policy. If you do nothing, then the economy will be stable. There are three useful pieces of valuable policy advice here: "Things are on track; let them be," "don't create more inflation in a boom with unnecessarily expansionary policy," and "try not to make a depression out of a recession with overly contractionary policy."

#### 14. What Is Good Monetary or Fiscal Policy?

There are important exceptions to the recommendation of neutral policy when the economy is at potential output. The first is when the economy is at potential output but inflation is higher than the Fed's desired inflation rate. For example, at the end of 2020, real GDP was \$19.81 trillion, while potential output was \$19.89 trillion. Inflation, however, was at 7.1%, much higher than the 2% rate that the Fed typically aims for. With inflation running so high, is the best monetary policy a neutral one? The best policy may turn out to be a contractionary policy, even if that puts the economy into a recession. The economy can eventually return to potential output but at the lower rate of inflation.

This type of policy action has an important precedent in US economic history: In 1979, Paul Volcker, the newly appointed chair of the Fed, clearly wanted inflation rates to be closer to 2% than the double-digit rates that had prevailed in the 1970s. Volcker pursued a contractionary monetary policy that resulted in a fairly substantial recession in the early 1980s in the United States. The short-term pain that the economy endured paid off: Inflation stayed in the 2% to 4% range for much of the next four decades. Thus, when inflation increased sharply in late 2020, some economists called for the Fed to announce a sharply contractionary policy to fight off the 7% inflation.

The second exception to the “leave well enough alone” rule is related to fiscal policy actions that are aimed at the long-term health of the economy. Over time, since output is greater than potential, inflation begins to rise. However, the end result of raising the level of potential output is a permanently higher level of output. There is still a higher inflation rate (but, again, not as high as it would be under an expansionary policy that did not raise potential output).

## Increasing the Potential Output Level

Increasing the potential output level of the economy can only happen if fiscal policy increases the economy's capital, labor, or technology levels. In theory, spending increases that raise potential output can include things such as more research spending into new technology development. Alternatively, tax cuts that provide more incentive to work and encourage firms to invest more in equipment and structures can also increase capital and labor in the economy and boost potential output.

#### 14. What Is Good Monetary or Fiscal Policy?

Whether such policies are implemented in the economy often leads to contentious debate. For example, consider the 2017 Tax Cut and Jobs Act that the Trump administration passed. The justification was that this bill represented the most significant corporate tax reform in many decades and would lead firms to move jobs back to the United States, invest in expanding their businesses, and produce more output. The expansionary impact of the policy would be to boost investment and increase the long-term potential output of the economy. Detractors argued that the bill's provisions to increase long-term potential output were too weak and that firms would mostly use the lower corporate taxes to increase pay for executives and boost their stock price through share buybacks.

In 2021, the Biden administration passed the \$1.3 trillion bipartisan American Jobs Act targeting the renewal of the nation's infrastructure. This bill arrived on the heels of more than \$4 trillion of expansionary fiscal policy from the \$2.2 trillion CARES Act signed by President Trump and the \$1.9 trillion American Rescue Plan signed by President Biden. This time, the proponents argued that the investment in infrastructure would strengthen the economy and increase its productive capacity in the long run. This claim was disputed by the bill's opponents, who focused on the inflationary consequences of passing such a bill when inflation was already rising rapidly. Note that this does not mean that all government purchases should be targeted toward things that can improve potential output. Increased research spending at universities or higher-quality early childhood education programs are probably not good recession-fighting tools—their effects will take too long to materialize.

The task of policymakers can be expressed in a succinct way: Move the economy to potential output and keep it there as long as the current inflation rate is acceptable. However, where exactly is potential output? Potential output is not something that can be measured directly. Consider what happened in the financial crisis of 2008/2009: When the recession hit, it became apparent that the impact would be deep and long-lasting. Labor markets in particular were hit hard, and the unemployment rate rose sharply. Unemployed workers seemed to be taking a long time finding jobs, and what seemed to be frictional unemployment was turning out to be structural unemployment. This led the Congressional Budget Office (CBO) to sharply lower its estimates of potential output. Thus, the long-run growth path itself can be influenced by the depth and severity of a recession. This in turn means that it is tricky to evaluate whether a policy response was a good one.

#### 14. What Is Good Monetary or Fiscal Policy?

There are also challenges posed by concurrent changes in the economy. For example, the US economy experienced high inflation in 2022. Some observers attributed that to policy mistakes by the Fed and the federal government in continuing to pursue expansionary policies even after the economy had recovered from the 2020 recession. Others attributed the inflation to supply shocks—some from the pandemic and others from the rapid rise in oil prices stemming from the war in Ukraine and the sanctions on Russia. The appropriate policy response will differ based on which of these explanations is more valid. If high inflation is the result of past mistakes in moving the AD curve out too much, the appropriate response will be to engineer a Volcker-style disinflation and enact a contractionary policy, even though the economy is currently close to potential output. However, if the high inflation is the result of temporary supply shocks stemming from war or the pandemic, then perhaps a neutral policy is called for.

### READING

Blinder, Alan, Jay Shambaugh, and David Wessel. “What Can US Fiscal and Monetary Policy Do to Limit the Economic Harm from COVID-19?” Brookings, March 10, 2020. <https://www.brookings.edu/blog/up-front/2020/03/10/what-can-u-s-fiscal-and-monetary-policy-do-to-limit-the-economic-harm-from-covid-19/>.

Congressional Budget Office. “A Summary of Alternative Methods for Estimating Potential Output.” Background Paper, March 2004. <https://www.cbo.gov/sites/default/files/108th-congress-2003-2004/reports/03-16-gdp.pdf>.

Elmendorf, Douglas, and Jason Furman. “If, When, How: A Primer on Fiscal Stimulus.” Hamilton Project Strategy Paper, January 2008.

Mallaby, Sebastian. *The Man Who Knew: The Life & Times of Alan Greenspan*. Penguin Books, 2016.

Wallace-Wells, Benjamin. “Larry Summers versus the Stimulus.” *The New Yorker*, March 18, 2021. <https://www.newyorker.com/news/annals-of-populism/larry-summers-versus-the-stimulus>.

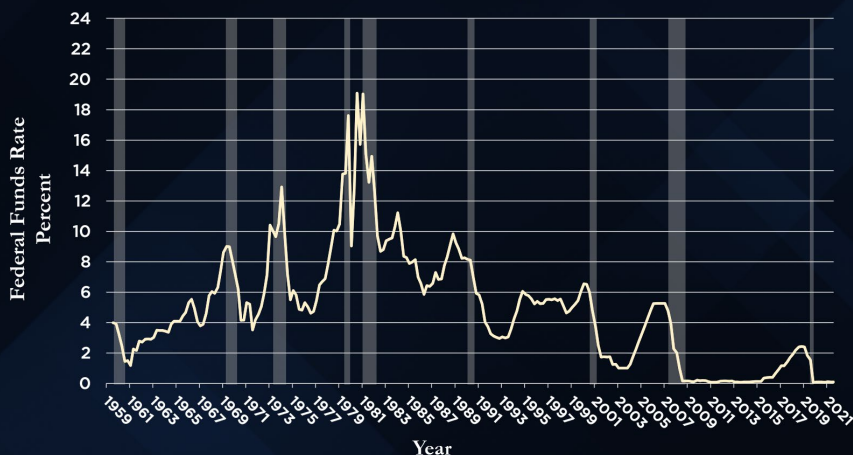
# 15

## EXTRAORDINARY ACTIONS BY CENTRAL BANKS

The primary tool that central bankers use to keep the economy at the “just right” state of affairs is the interest rate. However, sometimes a recession hits the economy so hard that policymakers cannot lower interest rates by enough to get it back on track. This is because interest rates have a lower bound of zero. Beginning with Japan’s experience in the 1990s, followed by the worldwide recessions of 2008/2009 and 2020/2021, central bankers have found the zero lower bound (ZLB) to be a real threat. This lecture examines the challenges posed by hitting the ZLB on interest rates and the policy options that central bankers can use in such circumstances.



Interest Rates in the US Economy from 1960–2021



## The Taylor Rule

When the economy is in an ordinary recession, the Fed typically acts aggressively to lower interest rates, raise spending, and get the economy back onto its growth path. Since 1960, the federal funds effective rate has dropped quickly and sharply whenever the country enters a recession. The more severe and prolonged the recession, the more the federal funds rate will fall.

To describe the interest-rate-setting behavior of the Fed, economists use the Taylor rule, credited to work published by the Stanford economist John Taylor in 1993. The Taylor rule describes how the Fed will typically respond to the state of GDP and inflation in the economy. It is a relationship that says the federal funds rate can be summarized as the sum of a desired nominal interest rate, an adjustment for inflation relative to a target rate, and another adjustment for the state of the economy: the gap between real GDP and potential output.

## 15. Extraordinary Actions by Central Banks

Essentially, if everything is on track, meaning inflation is at its target of 2% and the economy is on its growth path (real GDP equals potential output), the Fed will choose a nominal interest rate of 4% for the federal funds rate. Then, for every 1-percentage-point rise or fall in inflation, the Fed will raise or lower the nominal rate by 1.5 percentage points. For every 1 percentage point that real GDP rises or falls above its sustainable growth path, the Fed will respectively raise or lower the nominal interest rate by 0.5 percentage points.

Fed Funds Rate =

$$4\% + 1.5 \times (\text{Inflation} - 2\%) + 0.5 \times \left( \frac{\text{Real GDP} - \text{Potential Output}}{\text{Potential Output}} \right)$$

This simple relationship describes the behavior of the Fed from the mid-1980s through the 1990s and into the early 2000s surprisingly well. However, as inflation falls and the output gap (the difference between real and potential GDP) becomes more negative, the nominal interest rate suggested by this policy rule also falls. At some point, the nominal interest rate can hit the ZLB. This is what happened in 2008/2009 when the depth of the decline in real GDP and the fall in inflation resulted in the Taylor rule recommending that the Fed set a nominal interest rate of around -4%, which it couldn't do because of the ZLB.

## Understanding the ZLB

Why do interest rates have a ZLB? Suppose the interest rate was -2%. From a depositor's perspective, if you put \$100 into a bank account, at the end of one year, you will have \$98. From a lender's perspective, if you lend \$100 to someone, after one year, that person will have to pay you back \$98. In both cases, there is a superior alternative: Hold on to the money. The bank keeps the money in its vault and earns 0% rather than making the loan and earning -2%. The account holder keeps the cash and earns 0% rather than entrusting it to the bank and earning -2%.

In essence, when the ZLB is reached, the Fed can no longer rely on pushing the federal funds rate lower. The usual forces guiding the economy back from a recession—falling inflation leading to lower interest rates leading to more spending leading to higher GDP—are no longer operative. The economy can get caught in a liquidity trap—a deflationary recession that gets worse over time.

# Liquidity Traps

In a recession, because spending is much lower than the productive capacity of the economy, inflation falls. If the recession is severe enough, the economy can even move into deflation. Although falling inflation will usually lead the Fed to lower real interest rates and help boost spending in the economy, this becomes impossible to do at the ZLB.

Recall that the real interest rate is the nominal interest rate minus inflation. However, if the nominal interest rate is at its ZLB, the real inflation rate simply equals “minus inflation”—that is, the real interest rate moves in the opposite direction of inflation. When inflation rises, the real interest rate falls, and when inflation falls, the real interest rate rises. However, recall that in a deep recession, inflation falls, perhaps even turning into deflation. Thus, there is now a vicious cycle whereby in a recession, inflation falls, the real interest rate rises, and spending falls. This worsens the recession, resulting in even more of a fall in inflation, and so on. Breaking free of the liquidity trap requires that the Fed move to unconventional monetary policy actions.

In essence, not all interest rates hit zero when the federal funds rate hits the ZLB. Consider the interest rates (bond yields) of different maturities from 2006 to 2021. After the 2008/2009 recession, the federal funds rate and many shorter-term interest rates all went to zero—the rates for one month, three months, six months, and even 1 year. However, the 5-year rate did not go to zero, even though it did fall. Moreover, the rates for 10-year and 30-year bonds fell even less. This suggested the Fed still had scope for stimulating the economy if it could get those longer-term rates to drop.

To understand how (or if) the Fed can bring those long-term rates down, you need to develop a theory that relates how short-term and long-term interest rates are linked. You can derive a simple relationship between the yields of bonds of different maturities that will hold under certain assumptions.

## 15. Extraordinary Actions by Central Banks

Suppose you have \$100,000 that you want to invest for 10 years in a safe way. A risk-free way to do this would be to buy a 10-year US government bond today. Call the annual yield of that bond the “current yield of the 10-year bond.” You would then earn this yield for 10 years. This is investment strategy A. Alternatively, you could put the money into a 5-year US government bond. Call the annual yield of that bond the “current yield of the 5-year bond.” You would then earn this yield for the next 5 years. After the 5 years pass, you could buy another 5-year US government bond. Call the annual yield of that bond the “expected yield of the 5-year bond in 5 years’ time.” This is investment strategy B.

A simple theory of the relationship between long-term and short-term bonds, the expectations hypothesis, can be built on the notion that investment strategies A and B should generate the same end sum of money for you. The yield on today’s 10-year bond should be equal to the average of the yields of today’s 5-year bond and the expected yield of a 5-year bond 5 years from now. Mathematically, this translates into this equation: current 10-year bond yield = average of (current 5-year bond yield + expected 5-year bond yield in 5 years’ time).

Why does this relationship make sense? Both strategies are risk-free. Thus, if investment strategy A is more attractive, people will rush to buy 10-year bonds today. This will drive up the price of 10-year bonds, pushing 10-year bond yields down, and drive down the price of 5-year bonds, pushing 5-year bond yields higher. Remember, yields move in the opposite direction from prices. This will go on until the two strategies give the same return. The same story operates in reverse if investment strategy B is more attractive.

The expectations hypothesis, therefore, states that current long-term interest rates are equal to the average of current and future expected short-term interest rates. Basically, the relationship between short-term and long-term interest rates in the economy today reflects what people expect short-term rates to do in the future; that is, if short-term interest rates are expected to rise over time, then current long-term interest rates would exceed the current short-term interest rate. In contrast, if short-term interest rates are expected to fall over time, people would expect to find that current long-term interest rates are lower than current short-term interest rates.

## The Yield Curve

To see how this powerful idea plays out in real life, economists use a yield curve. This essentially plots interest rates (bond yields) of increasing term lengths. In typical periods of growth, the longer period of time has a higher rate. However, sometimes the yield curve is flat or even downward-sloping. That's what happened in December 2006, before the housing bubble collapsed and the banking failures of 2008 emerged. All of the yields were relatively close to one another; both short and long-term interest rates were around 5% in value. The slope of this yield curve is said to be flat.

How does the expectations hypothesis factor into this picture? It says, for example, that the six-month rate reflects the average of the current federal funds rate and expected federal funds rates over the next six months. Thus, this shows that in December 2006, people thought the Fed would keep interest rates relatively stable for the foreseeable future. The economy seemed to be in a good place—neither unusually hot (which would lead the Fed to raise the federal funds rate) nor unusually cold (which would lead the Fed to lower the federal funds rate). Now, fast-forward two years to December 2008.

Yield Curve: December 2006



## 15. Extraordinary Actions by Central Banks

Lehman Brothers had collapsed a few months prior. The housing bubble had popped, the recession had begun, and the panic was setting in. The federal funds rate had fallen to 1%; the one- and three-month rates were even lower than 1%. The yield curve had a clear upward slope to it. This was reflected in the unraveling of the tightly coiled rope of interest rates.

Why were the one- and three-month rates lower than the federal funds rate in December 2008? The expectations hypothesis says that the three-month rate, for example, reflects the average of the current federal funds rate and expected federal funds rates over the next three months. Thus, people were expecting the Fed to lower interest rates even further below the current level of 1% over the following three months. The fact that the one-year, two-year, and five-year rates were higher than the federal funds rate reflects the expectation in December 2008 that the then-low 1% federal funds rate would not last for much more than a year—rates would rise again.

Now, consider December 2010, by which time the reality of the Great Recession had settled in. The federal funds rate was at zero, and all the rates out to the two-year rate were at zero. For technical reasons, a zero interest rate is any rate that is less than 0.25%. Recall that when the federal funds rate becomes low in a time of ample reserves, the Fed can't influence it precisely

**Yield Curve: December 2008**



using open-market operations. The ZLB had been reached, and people were expecting it to stay for a while. Even the five-year rate was at 1%, indicating that a long period of lower interest rates was expected on the horizon.

How does this relate to monetary policy? Go back to December 2008 and suppose the Fed has already cut the federal funds rate to zero and can lower it no more. The Fed knows that the behavior of longer-term rates will depend not only on the value of the federal funds rate in December 2008 but also on what the expected value of the federal funds rate will be in the months ahead. Suppose the Fed announces that the federal funds rate will stay at zero over the next year. If people believe this, then all the rates up to one year will go to zero. Rates longer than one year—say, the five-year rate—will also fall but not necessarily to zero since the Fed may not commit to a path for policy that far into the future. However, since the current five-year rate is the average of current and expected future one-year rates, if the current one-year rate is driven down, that will also put downward pressure on the current five-year rate.

The key point is that the Fed can transmit its policy actions to the real economy in two ways. First, it can change the federal funds rate, which affects banks' lending behavior in the short run. Second, it can influence expectations about the future path of interest rate movements. These expectations impact longer-term bond yields. This can drive spending changes in the real economy, which depend more on long-term interest rates. This act of communicating information about the path of interest rates in the future is called forward guidance. However, it would be difficult for the Fed to bring down the 10-year and 30-year rates using forward guidance. Fortunately, there are other tools that the Fed can use, even when at the ZLB.

## Additional Fed Tools

The first option comes from the concept of a term premium. There is reason to believe that there is a natural tendency for long-term rates to be higher than short-term rates. Lenders may prefer short-term loans paying 5% interest to long-term loans paying 5% interest because they have less uncertainty about the macroeconomic circumstances such as inflation over short horizons. Meanwhile, borrowers would like to borrow at 5% over a longer term rather than 5% over a shorter term so that they do not face as much repayment pressure in the immediate future. The combination of a greater demand for

## 15. Extraordinary Actions by Central Banks

longer-term loans from borrowers and a greater supply of shorter-term loans by lenders results in a term premium: Borrowers have to pay a higher interest rate to borrow money for longer periods of time.

Thus, you can say the following: long-term interest rates = expected path of short-term interest rates + term premium. In terms of the yield curve, this means that the longer end (10 years, 30 years) will typically be higher than the shorter end, even when short-term rates are expected to stay constant over time. To reduce term premiums, which basically reflect the fact that bond buyers (lenders) are more reluctant to purchase longer-term bonds, the Fed can allocate more of its open-market operations toward those longer-term bonds.

After the 2008/2009 recession, even though the federal funds rate couldn't go any lower, the Fed decided to continue to increase the money supply in the economy. It used quantitative easing (QE) programs to inject large amounts of money into the economy through the purchase of the government's own long-term bonds. The QE2 program in 2010 saw the Fed buy \$600 billion worth of long-term Treasury securities. In 2011, Operation Twist had the Fed sell off some of its portfolio of short-term bonds and buy more long-term government bonds to bring down long-term interest rates in the economy. This was an attempt to twist the longer end of the yield curve down and shift the shorter end of the yield curve up.

The Fed has another powerful tool in its arsenal, which comes from the relationship between risky bonds and safe bonds. Recall that a risky bond will always sell for a lower price (meaning it has a higher yield) than a safe bond of the same face value. Since price and yield are inversely related, this means that risky bonds will always have a higher yield than safe bonds. Another way to state this is as follows: interest rate on risky assets = interest rate on safe assets of similar maturity + risk premium. The presence of risk premiums will present another opportunity for the Fed to boost the economy. To reduce risk premiums, the Fed can basically step in and buy assets that are riskier than US government bonds. At the depths of the financial crisis, the Fed vastly increased the types of assets it held, expanding especially into purchases of mortgage-backed securities (MBS), from \$0 in 2008 to more than \$1 trillion by 2010.

Fannie Mae and Freddie Mac are quasi-government agencies that buy home mortgages from banks and convert them into MBS. Sallie Mae does the same for student loans. Beginning in 2009, the Fed purchased bonds and MBS from Fannie Mae and Freddie Mac, student loans from Sallie Mae, and also

some high-quality corporate bonds. This way, it was able to lower riskier interest rates. This enabled consumers to access loans at a lower rate than they would otherwise have been able to.

During the 2020 recession, the Fed used the full panoply of conventional and unconventional policies discussed thus far. First, zero interest rates: The Fed acted swiftly in March 2020 and brought the federal funds rate to its ZLB. Second, forward guidance: In September 2020, the Fed said that rates would remain low until the economy was essentially back on track. Third, QE: In March 2020, the Fed initially bought \$500 billion in longer-term Treasury bonds and \$200 billion in MBS. Then, between mid-March and mid-June, the Fed increased its purchases of all securities by more than \$2 trillion to reduce term premiums on US Treasuries and risk premiums on MBS. It also restarted a 2008/2009 program to buy high-quality student loans, auto loans, and credit card loans with up to \$100 billion in new credit. Fourth, at the start of the pandemic, the Fed also supported at least nine programs aimed at short-term liquidity, including a Paycheck Protection Program. Think of this as reducing risk premiums in the economy. In 2020, purchases exceeded \$1.5 trillion at the peak of the COVID-19 recession.

### READING

Bernanke, Ben S. *21st Century Monetary Policy*. W. W. Norton, 2022.

———. *The Courage to Act: A Memoir of a Crisis and Its Aftermath*. W. W. Norton, 2017.

Gagnon, Joseph, Matthew Raskin, Julie Remache, and Brian Sack. “The Financial Market Effects of the Federal Reserve’s Large-Scale Asset Purchases.” *International Journal of Central Banking* 7, no. 1 (March 2011): 3–43.

Milstein, Eric, and David Wessel. “What Did the Fed Do in Response to the COVID-19 Crisis?” Brookings, December 17, 2021. <https://www.brookings.edu/research/fed-response-to-covid19/>.

Sorkin, Andrew Ross. *Too Big to Fail: The Inside Story of How Wall Street and Washington Fought to Save the Financial System—and Themselves*. Penguin, 2010.

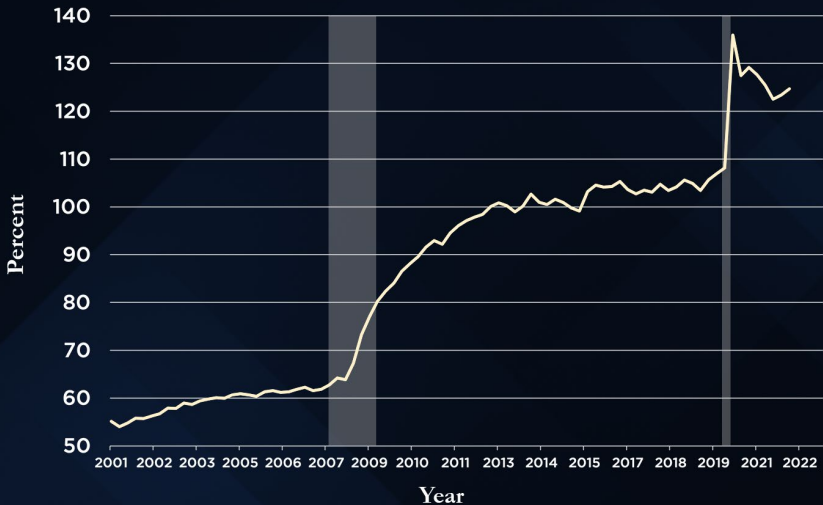
# 16

## HOW MUCH DEBT IS TOO MUCH FOR A COUNTRY?

**G**overnment debt is complex. Countercyclical fiscal policy can play an important role in stabilizing economic fluctuations. Running budget deficits and borrowing during bad times while running budget surpluses and paying off debt in good times is better policy than not taking on debt. However, as this lecture shows, how much debt a country should take on is a complex function of the currency it is borrowing in, what its debt-GDP ratio is, and how high the interest rate on the debt is relative to the growth rate of the economy.



Debt-GDP Ratio 2001–2021



## Deficits and Debt

By 2019, US debt had been rising for decades and had reached a plateau of about 100% of GDP, with a sharp increase following the recession in 2008/2009. Despite this, the average interest rate on the US debt had been steadily falling for those same two decades. The average cost of borrowing was around 6% in the early 2000s when the debt-GDP ratio was around 55%. However, by the 2020s, the cost of borrowing had fallen to under 2%, even though the debt-GDP ratio had more than doubled to more than 120%. Then, the world was hit by a global pandemic. Fiscal policymakers in the United States responded to the pandemic with extraordinary fiscal policy interventions that added more than \$3 trillion to government debt levels. Many other countries also increased their debt. Once inflation began rising in the economy, the Fed began raising short-term interest rates.

## 16. How Much Debt Is Too Much for a Country?

A budget deficit is the amount by which government expenditures, or government outlays, exceed government revenues. If it is zero, the budget is said to be balanced. If government revenues exceed government expenditures, then there is a budget surplus, or a negative budget deficit. Budget deficits are a flow variable: This variable reflects what is happening in the current fiscal year and does not say anything about what happened in the past. When a government runs a budget deficit, it has to borrow money to fund the difference between expenditure and revenue. The total amount of borrowed money that has not yet been repaid is the government debt. Debt is a stock variable: It reflects the cumulative impact of past and current actions. Basically, if a country runs a budget deficit, this adds to its debt. If a country runs a budget surplus, this reduces its debt.

The fiscal year for the US federal government runs from October 1 of the previous calendar year to September 30 of the current calendar year. Beginning in the late 1970s, with the exception of a brief period from 1999 to 2001, the federal government has consistently run budget deficits. The \$400 billion deficit in 2004 was less than 3.4% of GDP. Then, the 2008/2009 recession hit, and the country ran a \$1.4 trillion deficit in 2009, which was almost 10% of GDP. In 2020, the federal budget deficit was 15% of GDP, at \$3.1 trillion. This consistent flow of budget deficits implies that the stock of outstanding government debt has also been rising over the past 40 years. The total amount of government debt rose from \$290 billion in 1981 to \$29 trillion in 2021.

When the government releases data on the national debt, the most common series given is “Debt Held by the Public.” The distinction between the total amount of government debt and the debt held by the public is that some government bonds are held by the federal government itself—for example, in the Social Security trust fund. If the revenue from Social Security payroll taxes exceeds Social Security payments in a given year, the surplus is invested into government bonds. What about the large amount of US government securities that have also been purchased by the Fed in recent years? Those are counted as part of the debt held by the public since the Fed is an independent entity that buys and sells securities on the secondary market.

## 16. How Much Debt Is Too Much for a Country?

The typical convention for debt and deficit data is to express values as a percentage of GDP. That way, economists can gauge how large the borrowing needs or outstanding obligations are relative to the size of the economy. Prior to the 2008/2009 financial crisis, budget deficits had been in the range of around 2% to 5% of GDP, but they reached 10% of GDP in 2009 and 15% in 2020. The debt level held by the public fluctuated around 40% of GDP for much of the 1990s and 2000s. In the decade after the 2008/2009 financial crisis, the debt-GDP ratio gradually increased, reaching 80% of GDP before the pandemic onset. The unprecedented high deficits of 2020 resulted in the debt held by the public reaching 100% of GDP.

Another useful way to look at deficits is to look at the primary deficit, which takes out interest payments on debt. It shows how much of the deficit is due to current fiscal policy decisions as opposed to paying for the actions of past fiscal policy decisions. Budget deficits tend to worsen in bad times because social expenditures rise and tax revenues fall. Thus, economists will sometimes look at the structural deficit. This is the budget deficit that would prevail if tax revenues and social expenditures were more normal. The difference between the actual deficit and the structural deficit is the cyclical deficit. Changes in the structural deficit reflect longer-term budgetary misalignments that are sometimes dwarfed during a deep recession by cyclical factors.

The CBO produces 10-year forecasts of the budget deficit every year. For the United States, persistent budget deficits are predicted to be the norm for the 2020s. The primary reason is the aging of the US population. Beginning in 2021, current Social Security and Medicare tax revenues have not been sufficient to fund current Social Security and Medicare expenditures. That gap is forecast to grow to more than \$2 trillion from 2022 to 2031. Accordingly, the debt-GDP ratio will stay higher than 100% for the next decade.

## Implications

Borrowing costs may rise when a government takes on too much debt. If lenders worry that the obligations of a government are so large that it will not be able to honor future repayment, the price that they will be willing to pay for US government bonds will fall. Since yields and prices for bonds are inversely related, yields will rise. Effectively, the US government will have to pay more to borrow money. With debt-GDP levels of around 100%, interest

## 16. How Much Debt Is Too Much for a Country?

payments of  $x\%$  on the debt will effectively mean a budget deficit of  $x\%$  even if the primary budget balance is zero—that is, even if expenditures outside of interest payments match tax revenues.

The aging of the US population will put rising pressure on budget deficits as Social Security and Medicare payments outstrip payroll tax revenues. If interest rates rise, budget deficits will rise, making the debt-GDP ratio worse and causing rates to rise even more. Moreover, too much debt will have damaging effects on the economy far beyond this. Recall also that the interest rates on US government bonds drive other interest rates in the economy. As other borrowing rates increase, investment in the domestic economy will fall. This is called crowding out, and it has adverse long-run consequences for the economy because less capital is being built for the future. If GDP falls, the debt-GDP ratio will rise, further increasing uncertainty about debt sustainability and worsening the vicious cycle.

At some point, the government will have to make spending cuts and increase taxes. This could harm growth, worsen poverty, or increase inequality. If imbalances lead the government to reduce Social Security payments or delay the age at which people can claim benefits, this will disproportionately impact people who have no other means of income support in their old age. The same goes for Medicare. Cutbacks in payments, increased eligibility ages, and higher premiums will all have a greater impact on the segment of the population that is unable to afford private insurance costs after retirement.

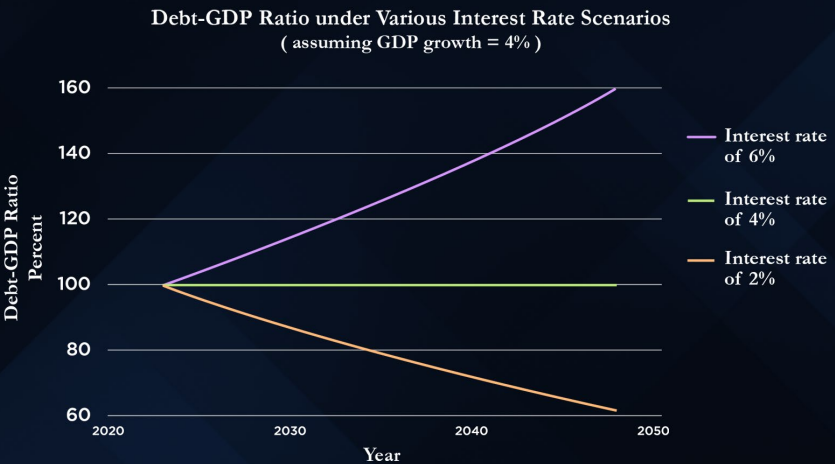
There is also the issue of intergenerational fairness. If the government spends more and taxes less now, borrowing to make up the difference, that is a double benefit to the current generation: They get to enjoy the fruits of the greater spending and also get to keep more of their current income thanks to the lower taxes. However, paying off the debt will require that future governments spend less and tax more. This will be a double burden on future generations.

All of these problems rest on the premise that interest rates will rise as debt-GDP ratios increase. Clearly, at some point, the chain of events that links higher levels of debt to greater uncertainty to higher interest rates will hold. However, at what point do investors begin to doubt the ability of the government to repay its promises? Consider the relationship between the level of government debt and the primary budget deficit (the budget balance

## 16. How Much Debt Is Too Much for a Country?

excluding interest payments on the debt). The debt in each successive year is equal to the following: debt this year = last year's debt + interest + the primary deficit this year. Now, divide through by GDP to get the debt-GDP ratio.

Suppose that GDP increases at a rate of  $g$ . Then, you can express the debt-GDP ratio as being related to last year's ratio by a factor that depends on the ratio of the interest rate on the debt to the growth rate of GDP. Now, suppose there is a situation where the debt-GDP ratio is 100% and the government is committed to keeping the primary budget deficit at zero. Suppose that nominal GDP is growing at 4% (i.e.,  $g = 4\%$ ). Consider three alternative situations where the average cost of borrowing ( $i$ ) is equal to 2% or 4% or 6%. If the interest rate on the debt stays at 2%, which is lower than the 4% growth rate of GDP, then the debt-GDP ratio falls nicely over time, even without running a primary budget surplus. However, if the interest rate on the debt is higher than GDP growth, then even if the primary balance is zero, the debt-GDP ratio rises over time to 160%.



Thus, even though the US debt-GDP ratio in 2019 had risen for a decade, the fact that interest rates on the debt were so low meant that it was not as threatening to have a large debt-GDP ratio. The United States has since

experienced the pandemic and the resulting large deficits in 2020 and 2021 that caused the debt to rise dramatically. Interest rates on the debt have remained low. However, as inflation is rising and the Fed is raising interest rates, the possibility that the growth rate of GDP will fall and interest rates on government debt will rise means that the future of debt dynamics may not be so stable. Thus far, economists have not identified a specific value of debt, relative to the size of the economy, at which people lose confidence in the ability of a government to repay its debts. As such, governments have to make sure that they pursue sensible policies that keep debt under control.

## Safe Assets

One argument for why US government bonds have retained their high prices and low yields in the midst of this vast expansion of government borrowing concerns rising global demand for safe assets. In an influential paper in 2006, Ricardo Caballero, Emmanuel Farhi, and Pierre-Olivier Gourinchas argued that the expanding global economy in the 21st century increased demand for safe assets—particularly to store the vast surpluses and foreign reserves in China and in oil-exporting nations.

At first, this demand was filled by creating assets that seemed ostensibly safe: MBS and derivatives in the United States, bonds issued by eurozone countries such as Greece and Portugal, and short-term securities issued by US investment banks. However, the 2008/2009 financial crisis showed that these assets were far from safe. The US policy response was an increase in borrowing by the US federal government of almost \$6 trillion in the five years after the financial crisis. The evaporation of confidence in other financial assets meant that the demand for safe assets rose even more dramatically than the supply did, driving US bond prices higher and yields lower. The purchase of US government bonds by the Fed is also likely to have played a role in keeping bond yields low in the face of increased supply.

An even more extreme example of asset purchases by a central bank comes from Japan. Japan hit the ZLB a decade before the rest of the world. It experienced years of economic stagnation, repeated attempts to boost the economy using fiscal policy, and soaring debt-GDP ratios long before the pandemic. This is how Japan's debt-GDP ratio in 2020 rose to 250%. Yet, the interest rate on Japanese government bonds remained close to zero. This was because the asset

purchases by Japan's central bank were also more extreme. The Bank of Japan was buying so many government securities that the central bank had effectively become the only client. There was virtually no open trading of Japanese government bonds in the bond market anymore. Thus, it is unclear whether Japanese bonds are risk-free, because the entity (the Bank of Japan) that can costlessly print Japanese yen keeps buying all of those bonds.

# Constraints for Foreign Countries

Other countries should be careful about modeling their own borrowing behavior on the United States. Brazil may have a debt-GDP ratio of 90%, but the Brazilian government's promises to repay come in the Brazilian currency, the real. Although it is just as certain that the Brazilian government will make the promised future payment in real as it is that the US government will make its promised future payment in dollars, the dollar is valued more because of the likelihood that it will continue to be a more valuable medium of exchange in the future. Thus, countries without a strong currency are likely to face much more binding constraints on fiscal policy and unfavorable debt dynamics at a debt-GDP ratio much lower than the United States or Japan.

These countries will often try to borrow in dollars by issuing international sovereign bonds. Although they may be able to borrow at lower costs by promising future repayment in dollars, the constraint on how much foreign currency denominated debt a country can issue is likely to be much more binding. When debt levels become high and a lot depends on how much confidence potential lenders have in your government's ability to repay loans, standard fiscal policy actions become problematic. That's when counterintuitive policies come to the fore.

Consider the situation in Greece after the global financial crisis, when debt-GDP ratios had gotten too high. In Greece, the rising debt levels meant that expansionary fiscal policy would be challenging to implement because it would require more borrowing. Moreover, there was a reluctance to lend to Greece because of the perceived risk. Greece also had no control over its currency because it was part of the eurozone. Therefore, there was no confidence in the government's ability to repay. Rising interest rates on Greece's government bonds further reduced confidence. Higher interest

## 16. How Much Debt Is Too Much for a Country?

rates then spilled over to other Greek borrowers, and consumption and investor spending fell. The reduction in consumption and investment would potentially overwhelm and reverse the effects of any government spending.

This sometimes leads to a suggestion that a country like Greece should counterintuitively follow a contractionary fiscal policy during a recession. The argument is that by reducing the budget deficit and, thus, reducing the rise of government debt, policymakers will create more confidence among investors. This greater confidence will supposedly lead to higher demand for government bonds, driving yields lower. Lower yields on government bonds will spill over into other interest rates in the economy, leading to more consumption and investment—more than offsetting the fall in government expenditures.

However, asking Greece to follow such a contractionary policy when it is already facing a deep recession is putting too much faith in investors—in particular, their ability to be convinced that the policy would increase the likelihood that the Greek government will repay its obligations. In fact, even though Greece ran substantial primary budget surpluses from 2011 to 2016, the interest rate on Greek bonds remained high. As such, the overall deficits also remained high, which caused the debt-GDP ratio to grow worse.

### READING

Blanchard, Olivier. “Public Debt and Low Interest Rates.” AEA Presidential Address, 2019. <https://www.piie.com/commentary/speeches-papers/public-debt-and-low-interest-rates>.

Graeber, David. *Debt: The First Five Thousand Years*. Melville House, 2011.

McBride, James, and Anshu Siripurapu. “The National Debt Dilemma.” Council on Foreign Relations. Accessed December 12, 2022. <https://www.cfr.org/background/national-debt-dilemma>.

Papaconstantinou, George. *Game Over: The Inside Story of the Greek Crisis*. CreateSpace Independent Publishing Platform, 2016.

Reinhart, Carmen M., and Kenneth S. Rogoff. *This Time Is Different*. Princeton University Press, 2009.

# 17

## HOW EXCHANGE RATES MOVE THE GLOBAL ECONOMY

**H**aving different currencies can complicate transactions across economies. If you want to buy Swiss chocolates, your decision is influenced not only by the price of Swiss chocolates in Swiss francs but also by how many dollars it takes to acquire the Swiss francs needed to buy the chocolate. The price at which one country's currency can be exchanged for another plays an outsized role in influencing patterns of trade, investment, remittances, and migration. Thus, this lecture aims to help you understand how exchange rates are determined, why they are so volatile, and how they are influenced by various macroeconomic variables.

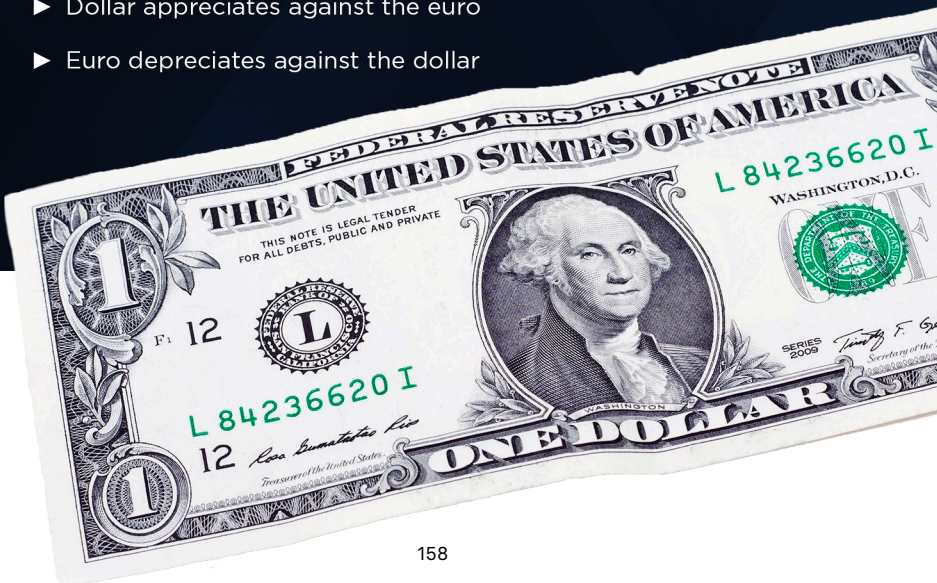


## What Is an Exchange Rate?

An exchange rate is the rate at which one country's currency can be exchanged for another country's currency. The clearest way to think about exchange rates is as the price of foreign currency: How many units of your money does it take to buy a unit of foreign currency? If the dollar increases in value against another currency, the dollar has appreciated. When the dollar appreciates, it may take only \$1.15 to buy a euro instead of \$1.25, for example. Note that if the dollar appreciates against the euro, then that is equivalent to the euro depreciating against the dollar. If the dollar falls in value against another currency, the dollar has depreciated. When the dollar depreciates, the exchange rate rises. It may now take \$1.35 to buy a euro instead of \$1.25. If the dollar depreciates against the euro, then that is equivalent to the euro appreciating against the dollar.

### Exchange Rate Changes from \$1.25/Euro to \$1.15/Euro

- ▶ Dollar has become more valuable against the euro
- ▶ Dollar appreciates against the euro
- ▶ Euro depreciates against the dollar



## 17. How Exchange Rates Move the Global Economy

You can use the exchange rate to compare the prices of domestic goods and foreign goods. Suppose a young German law student wants to buy a \$2,500 MacBook Pro. From her perspective, the \$1.25 per euro exchange rate is €0.8 per dollar. The MacBook costs €2,000. However, when the euro depreciates (the dollar appreciates) to €1 per dollar, the MacBook's cost increases to €2,500. If the euro appreciates (the dollar depreciates) to €0.667 per dollar, the price of the MacBook falls to €1,333.33.

A depreciation of the domestic currency makes foreign goods more expensive at home and domestic goods cheaper abroad. This can impact trade patterns. All else equal, depreciation leads an economy to export more and import less. Appreciation leads an economy to import more and export less.

For many countries, the price of a currency is determined by supply and demand conditions in the foreign exchange market, which is a gigantic worldwide electronic trading system of currencies. Participants in the market include commercial banks and investment banks performing transactions for corporations and individuals plus their own trading. Participants also include large companies conducting routine transactions worldwide; mutual funds, pension funds, and hedge funds that invest in stock markets and lend money to foreign companies; and central banks.

If the demand for foreign currency rises or the supply of foreign currency falls, the price of foreign currency rises (the domestic currency depreciates). Conversely, if the demand for foreign currency falls or the supply of foreign currency rises, the price of foreign currency falls (the domestic currency appreciates). Note that factors affecting demand and supply are likely to overlap fairly heavily in the foreign exchange market. Any economic event that raises the demand for foreign currency is likely to lower the supply of foreign currency as well.

This combined impact of supply and demand factors is seen in other financial markets as well. For example, if Tesla releases a good earnings report, the demand for Tesla shares will rise, but the supply will fall since existing shareholders will be less willing to part with their shares. The combination of a demand increase and a supply decrease (or a demand decrease and a supply increase) means that prices will move much more sharply up (or down) in currency markets and other financial markets than in goods markets.

## Absolute PPP

One of the most widely used theories of long-run exchange rate behavior is called PPP. Economists consider PPP to come in two versions. Absolute PPP says that in the long run, the price levels of goods and services in two countries, when measured in the same currency, should be identical. Mathematically, absolute PPP can be expressed as  $P = eP^*$ , where  $P$  is the domestic currency price of goods and services in the domestic economy,  $P^*$  is the foreign currency price of goods and services in the foreign economy, and  $e$  is the exchange rate. The left side is the domestic currency price of domestic goods and services, while the right side is the domestic currency price of foreign goods and services.

The intuition underlying this theory is that if prices of goods and services in the domestic economy are substantially higher than in a foreign country ( $P > eP^*$ ), over time, it is expected that the demand for foreign goods and services will rise (causing  $P^*$  to rise). Then, the demand for foreign currency to use to buy those goods and services will also rise, causing the dollar to depreciate ( $e$  will rise). This will ensure that the prices of goods in the two countries become more equal over time, as both  $P^*$  and  $e$  will rise until  $P = eP^*$ .

If domestic goods and services are more expensive, if  $P > eP^*$

Demand for foreign goods rises ►  $P^* \uparrow$  and

Demand for foreign currency rises ►  $e \uparrow$  until

$$P = eP^*$$

The argument works in reverse if prices of goods and services in the domestic economy are substantially lower than those in a foreign country. In that case, over time, the demand for domestic goods and services will rise (causing  $P$  to rise). The demand for domestic currency to use to buy those goods and services will also rise, causing the domestic currency to appreciate. This will ensure that the prices of goods in the two countries become more equal over time, as  $P$  will rise and  $e$  will fall until  $P = eP^*$ . Thus, the theory of absolute PPP says that in the long run, the exchange rate of an economy will equal the ratio of price levels

## 17. How Exchange Rates Move the Global Economy

in the domestic and foreign economies. This exchange rate is also called the PPP exchange rate. It can be used to calculate the PPP-adjusted GDP measure, which is the best measure to compare GDP across countries.

However, there are some serious problems with this theory. First, to calculate this PPP exchange rate, you need measures of the price level in the two countries. Take the most widely used measure of the price level for an entire economy: the CPI. The PPP exchange rate is expressed as the ratio of the domestic CPI basket (in domestic currency terms) to the foreign CPI basket (in foreign currency terms). However, a significant problem with the CPI is that these baskets contain different goods and services.

One way of avoiding this issue is to construct the PPP exchange rate with an identical basket of goods and services and measure the price of that basket across different countries. The International Comparisons Project run by the United Nations and the University of Pennsylvania constructs such a basket. It calculates the price of a basket of goods and services in more than 150 countries and determines a PPP exchange rate for each country as the ratio of the cost of the basket in the domestic and foreign economies. However, do not expect that the actual exchange rate will move toward the PPP exchange rate. Transportation costs and trade barriers can effectively drive a wedge between the price of goods across two countries.

Such disparities are further compounded for services. Suppose that the services of a home health care worker can be hired for \$20 per hour in the United States and the equivalent of \$15 per hour in Ireland. Someone who needs daily care in the United States can't hire the cheaper health care service in Ireland. Unless care providers can migrate without limit into the United States, you would not expect those prices to equalize over time. In real life, the absolute PPP theory that two countries will have the same price level for goods and services because of exchange rate movements is unlikely to hold true. That said, you can still use the PPP exchange rate to compare PPP-adjusted GDP numbers across countries.

## PPP-Adjusted GDP

PPP-adjusted GDP asks how two countries' GDP values would compare when measured using a common set of prices. The problem with absolute PPP is in asserting that the two countries would have the same price level for goods and

services because of exchange rate movements. Thus, what can be said about determinants of the exchange rate in the long run? A weaker version of PPP that is known as relative PPP is needed. In other words, absolute PPP is unlikely to hold because prices of goods and services can differ across countries due to trade and transportation barriers. However, once those differences are accounted for, the change in the price of foreign goods to domestic consumers should equal the change in the price of domestic goods to domestic consumers.

The increase in the price of foreign goods to domestic consumers depends on both the foreign inflation rate (i.e., how much prices of foreign goods are changing in foreign currency terms) and the depreciation of the domestic currency (i.e., how much more of it is needed to buy a unit of foreign currency). If you rearrange this equation, you get the key prediction of relative PPP: The country with a higher rate of inflation will see its currency depreciate against the currency of a country with a lower rate of inflation. Countries with higher inflation rates than the United States will see their currency lose value against the US dollar. Conversely, countries with lower inflation rates than the United States will see their currency gain value against the US dollar. The theory also seems to broadly hold true for countries with inflation rates comparable to the United States, such as New Zealand.

Relative PPP rests on the assumption of changes in trade patterns and the movement of goods and services, and that takes time. Thus, it is not a useful theory about short-term movements in the exchange rate. Such movements are more likely to be related to the movements of money across countries in the foreign exchange market. Flows of money across countries will be related to the rates of return available in those countries. In short, a theory of short-term fluctuations of exchange rates is more likely to be related to interest rates than to inflation. This theory is called interest rate parity. In its simplest form, it asserts that the expected rate of return from investing in the domestic economy equals the expected rate of return from investing abroad.

Take the United States and Turkey as examples. If a US investor invests in the United States, then she earns the interest rate in the United States on each dollar she has. Alternatively, she can take that money and invest it in Turkey. To do that, she has to first convert her dollars into Turkish lira and then earn the Turkish interest rate before converting her money back to dollars in the future. Therefore, if the US interest rate rises, US investors will prefer to keep

## 17. How Exchange Rates Move the Global Economy

their money in the United States. Turkish investors will prefer to convert their money into US dollars and earn a higher return. Thus, there will be an immediate appreciation of the dollar. However, if the Turkish interest rate rises, Turkish investors would prefer to keep their money in Turkey. US investors will move their money to Turkey, causing the Turkish currency to appreciate (which is the same as the dollar immediately depreciating).

Thus, in the short run, the dollar will appreciate when US interest rates rise (or foreign interest rates fall). Conversely, the dollar will depreciate when US interest rates fall (or foreign interest rates rise). For example, when US interest rates rise above UK interest rates, the pound depreciates against the dollar (the dollar appreciates against the pound).

## The Forward Market

People who know that they need access to a particular quantity of a foreign currency at some point in the future must protect themselves against unexpected exchange rate movements. Consider the following example of the owner of a specialty food store in the United States who wants to import 5 million cans of mushy peas, each costing £1, from England. Payment of £5 million is due in 90 days. The current spot rate is \$2 per pound; thus, currently, she is looking at a \$10 million cost. Exchange rate fluctuations, however, could change the cost to the importer by the time the payment date rolls around: If the dollar depreciates from \$2 per pound to \$2.25 per pound, then the cost increases from \$10 million to \$11.25 million. The importer is then looking for ways to protect herself against the depreciation of her currency. In international finance, this is called hedging.

	Exchange Rate	Expected Cost
Current	\$2.00 = £1	\$10 million
In 90 Days	\$2.25 = £1	\$11.25 million

The importer can hedge against fluctuations by buying the foreign currency right now on the foreign exchange market. However, this can be costly for two reasons: First, she may need to sell the goods she imported to obtain enough money to pay for the goods she bought. If she does not have the money up front, she has to borrow money. Second, even if she does not need to borrow, if she buys foreign currency now, she forgoes the opportunity to leave her money in the bank earning interest. In either case, acquiring the pounds today and paying off the foreign supplier carries a cost. As a result, many would-be hedgers have a keen interest in finding a hedging mechanism that does not require them to tie up their money in foreign currency up front yet protects them against fluctuations. The forward market provides them with that ability.

Think of that US importer just described—worried about how to protect herself against the depreciation of the dollar. There may be another trader in the United Kingdom who is also worrying because he has bought some goods from the United States and needs a certain amount of dollars in the near future. He is worried about the possibility of the pound depreciating. The place where both parties can come together and resolve their uncertainty is the forward market. In that market, an exchange rate is agreed on now for delivery of the foreign currency at some point in the future. The forward exchange rate is simply the rate that equilibrates supply and demand from these groups in the forward market.

The forward market is a great facilitator of hedging. Stay with the example of a specialty food store in the United States that needs to acquire £5 million in 90 days. Suppose that the forward rate is \$1.80 per pound. The importer can protect herself against exchange rate depreciation by buying £5 million for a cost of \$9 million on the forward market. Over the 90-day period, even if the dollar depreciates to \$2.50 per pound, her cost remains at \$9 million: She has successfully hedged against this downside risk. Note that hedging on the forward market does not require her to put the \$9 million up front—only the price that is agreed on today. Thus, forward markets provide a way to insulate traders from the volatility of exchange rates.

## READING

BBC. “Exchange Rate Roulette: Why It’s So Hard to Predict International Money Market Movements.” Accessed December 12, 2022. <https://www.bbc.com/storyworks/capital/a-world-of-opportunities/exchange-rate-roulette>.

Eichengreen, Barry. *Globalizing Capital*. Princeton University Press, 2019.

Magnusson, Gylfi. “What’s the Lesson of Iceland’s Collapse?” *Yale Insights*, May 22, 2009. <https://insights.som.yale.edu/insights/whats-the-lesson-of-icelands-collapse>.

The Economist. “The Big Mac Index.” July 21, 2022. <https://www.economist.com/big-mac-index>.

Williamson, John. “Exchange Rate Economics.” *Open Economies Review* 20, no. 1 (2009): 123–146.

# 18

## WHY COUNTRIES TRY TO CONTROL EXCHANGE RATES

In July 1944, delegates from 44 nations, including John Maynard Keynes, gathered for the Bretton Woods Conference at the Mount Washington Hotel. At this conference, the delegates established the IMF. They agreed to set up a system of fixed exchange rates. Signatory countries would fix the value of their currencies to the dollar, and the dollar would be convertible at a fixed rate into gold. This system collapsed in the early 1970s. Even then, many countries maintained their own fixed exchange rate systems. This lecture takes a closer look at how countries fix their exchange rates and why they might choose to do so.



## Fixing Currency

Under a fixed exchange rate system, the central bank announces a fixed exchange rate for its currency against a particular foreign currency, believing that a stable exchange rate will help facilitate trade and investment flows between countries by reducing fluctuations in relative prices and uncertainty. When a country chooses to fix its currency, it will typically choose to fix to the currency of a large trading partner or an important source of foreign investment.

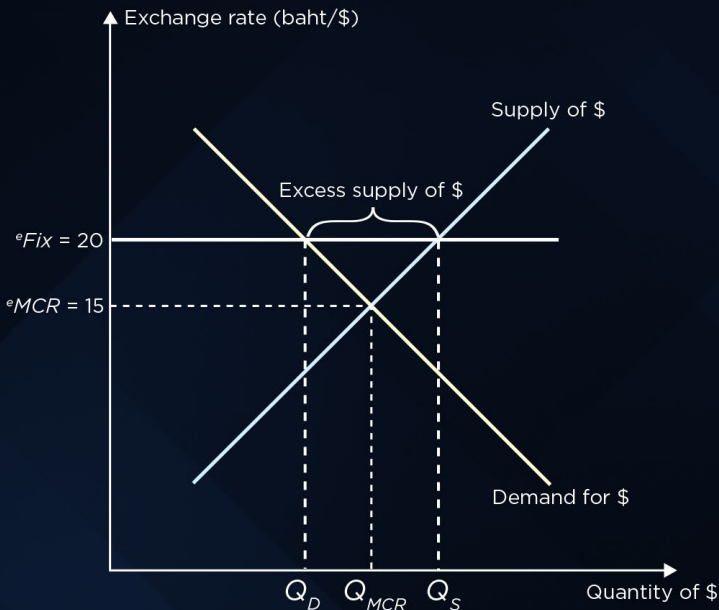
The IMF releases a report every year that assesses how each country sets the value of its currency. As of 2021, the IMF categorized 63 countries as having a market-determined floating exchange rate, of which 31 (including the United States, the United Kingdom, the eurozone nations, Japan, and Canada) were categorized as free-floating, which means there is little to no central bank intervention in influencing prices in currency markets. Simultaneously, a much larger group of around 100 countries have tied the value of their currency to a larger economy. For example, oil exporters and Caribbean nations who rely on US tourism fix to the dollar.

In a fixed exchange rate system, the central bank stands ready to exchange, at a pre-announced set rate, local currency and the foreign currency it has chosen to fix to. To support a fixed exchange rate, the central bank needs to have stocks of domestic and stocks of foreign currency (or foreign currency reserves). Typically, most foreign exchange reserves are held in terms of the currency that the country is fixing its currency to. For example, suppose that the Thai central bank has chosen to fix the value of its currency to the US dollar at 20 baht per dollar. Thus, for every dollar you'd bring to the Thai central bank, the central bank would give you 20 baht, and vice versa.

However, even though the central bank is announcing a fixed exchange rate, there is an implicit market clearing rate. This is the rate at which the quantity of foreign currency demanded by everyone other than the central bank equals the quantity of foreign currency supplied by everyone other than the central bank. However, foreign exchange markets are volatile. Supply and demand conditions change rapidly. The fixed rate will typically not be identical to this market clearing exchange rate. Depending on where the fixed exchange rate is relative to the market clearing rate, there will be situations of excess demand for or excess supply of foreign currency.

# Undervaluation and Overvaluation

Suppose that the Thai central bank chooses to fix the value of its currency to the US dollar at a rate of 20 baht per dollar when the market clearing exchange rate in Thailand is 15 baht per dollar. Effectively, the Thai central bank is making the baht artificially weak. Thus, investors will rush to move their dollars to the central bank. Since the fixed exchange rate makes the baht artificially less valuable, the baht is currently undervalued. This undervaluation creates an excess supply of dollars. At the market clearing rate of 15 baht per dollar, the supply and demand for dollars are equal. However, at the fixed rate of 20 baht per dollar, the supply of dollars exceeds the demand for dollars, meaning there is an excess supply.



## 18. Why Countries Try to Control Exchange Rates

What are the consequences of this undervaluation? The central bank absorbs the excess supply of dollars by selling baht in exchange for dollars. In the process, it acquires reserves of US dollars and runs down its baht reserves. Since the baht is undervalued, Thai exporters can sell more of their goods, and foreign investors can get more baht for their dollars. However, importers based in Thailand and Thai consumers seeking goods from abroad are forced to pay more. Thai entrepreneurs trying to invest in foreign economies also get fewer dollars for their baht.

However, exchange rates can change on a dime. Suppose the market clearing rate changes to 25 baht per dollar. Since the exchange rate is still fixed at 20 baht per dollar, the fixed exchange rate now makes the baht artificially more valuable. Now, the Thai central bank is offering a dollar for only 20 baht, when the market values a dollar at 25 baht. Thus, there is an excess demand for dollars. Since the fixed exchange rate makes the baht artificially more valuable, the baht is currently overvalued. This creates an excess demand for dollars. At the market clearing rate of 25 baht per dollar, the supply and demand for dollars are equal. However, at the fixed rate of 20 baht per dollar, the demand for dollars exceeds the supply of dollars, meaning there is an excess demand.

What are the consequences of this overvaluation? The central bank absorbs the excess demand for dollars by selling dollars in exchange for baht. In the process, it runs down reserves of US dollars and increases its baht reserves. Now, Thai exporters trying to sell goods to foreign countries and foreign entrepreneurs in Thailand get fewer baht for their dollars. Thai importers and consumers who need access to cheap goods from abroad and Thai entrepreneurs trying to invest in foreign economies get more dollars for their baht.

In an ideal world, a country's fixed exchange rate would not be systematically overvalued or undervalued. Periods of undervaluation would be interspersed with periods of overvaluation. This would result in greater exchange stability and no systematic benefit to one group of the population or to one sector of the economy. The central bank would periodically run down its foreign reserves and periodically build them back up again. In practice, though, this is not common. Governments will either choose fixed exchange rates that systematically undervalue or overvalue their currencies or end up in situations where their currency is systematically overvalued or undervalued;

that is, countries that fix their exchange rate will tend to face an extended period of time where the market clearing rate and the fixed rate diverge from each other.

A government may choose to systematically overvalue its currency to access imported resources or placate consumers with cheap access to foreign goods. It may also choose to do so if powerfully connected people control import licenses in the economy. However, some governments choose to systematically undervalue their currency to try and boost their economy through higher exports or attract increased foreign investment. Japan, for example, kept the yen artificially undervalued for more than a decade after the end of Bretton Woods to support exports. In 1985, Japan agreed to let the yen value appreciate more fully, at which point it appreciated from 239 yen per dollar in 1985 to 128 yen per dollar in 1988.

## Sustained Overvaluation

Consider what happens in a scenario where overvaluation persists over time. The central bank will continue to run down its foreign exchange reserves and correspondingly increase its domestic currency holdings. Having too few foreign exchange reserves is a serious problem. Suppose the excess demand is \$10 billion per month and the Thai central bank has \$100 billion in reserves. In 10 months, the reserves will run out. When the central bank runs out of reserves, it can't honor the fixed exchange rate anymore, meaning the fixed rate can't be sustained.

When the central bank does not have enough foreign exchange reserves, it must do one of three things: First, acquire enough foreign exchange reserves to convince traders that the current regime can be sustained. Second, change the fixed rate to make the currency less overvalued. This is called a devaluation: The domestic currency becomes less valuable. For example, in 1984, the Thai central bank moved from 20 baht per dollar to about 25 baht per dollar. Third, abandon the fixed exchange rate regime entirely and let the currency float. This means allowing the exchange rate to be determined in the market. This was what Thailand did in 1997 during the Asian financial crisis. The baht was allowed to float, and it soon took twice as many baht to buy dollars as before.

## 18. Why Countries Try to Control Exchange Rates

How would Thailand choose between the three options? Suppose that a cyclone damages the rice crop in Thailand. This creates an excess demand for dollars in Thailand because falling rice exports bring in fewer dollars. Since this is a natural disaster, Thailand may be justified in borrowing some reserves. However, if the demand for Thai rice exports is falling because of the entry of a new competitor into the global rice market, then the shock will not go away after a year. Thus, Thailand will have to make its rice prices more competitive via a devaluation of its currency. Finally, if declining soil conditions in Thailand or a global shift by consumers toward less white rice consumption causes the decline in rice exports, then even a one-time devaluation may not be sufficient. Thailand will have to let its currency float.

Now, suppose that you knew that the Thai government would be forced to devalue from 20 to 25 baht per dollar. If you had money in Thailand and knew that the exchange rate would jump next month, you would try to get your money out of Thailand. If you had 2 million baht, you could convert it into \$100,000 and then convert it back to 2.5 million baht after the exchange rate changed, earning a 25% rate of return. Many people would be rushing to convert baht into dollars if it looked like the central bank was running into difficulty with its foreign exchange reserves.

Conversely, if you had dollars and were looking to convert them into baht, you would wait until the central bank abandoned the fixed exchange rate. Then, you could get 25 baht per dollar. This implies that there will be few people wanting to convert dollars into baht when the central bank's reserves run low. The combination of a higher demand and a lower supply of dollars means that the excess demand for dollars at the fixed exchange rate would be even greater. Thus, the Thai central bank would run out of dollar reserves faster. This would lead to a dramatic rise in the number of people wanting to convert baht into dollars, and soon, panic would set in. If the participants in the foreign exchange market believe the central bank doesn't have enough reserves, that usually means it does not have enough reserves.

Yelling "devaluation" in a fixed exchange rate regime causes a melee that leaves people bruised, battered, and in pain. Furthermore, once doubts have set in, it's unlikely that a single devaluation or a single acquisition of reserves will be sufficient. People will continue to take more money out of the country. The country's central bank may end up borrowing more or devaluing more

than had initially been planned on. The impact of devaluation can also spill over to neighboring countries as panic spreads. For instance, in the Asian financial crisis, a sudden devaluation of the Thai baht soon had devaluating effects on other currencies: The Malaysian ringgit fell by 45%, the Korean won fell by almost 50%, and the Indonesian rupiah dropped by 80%.

# Sustained Undervaluation

Suppose that there is undervaluation that persists over time. The central bank will continue to accumulate foreign exchange reserves. In this case, the fixed exchange rate regime does not need to be abandoned. Holdings of domestic currency are declining, but this is not an immediate problem because the central bank can always print more of the domestic currency. Thus, undervaluation is not the mirror image of overvaluation. There is no similar imperative need for the central bank to act if the undervaluation gets too large.

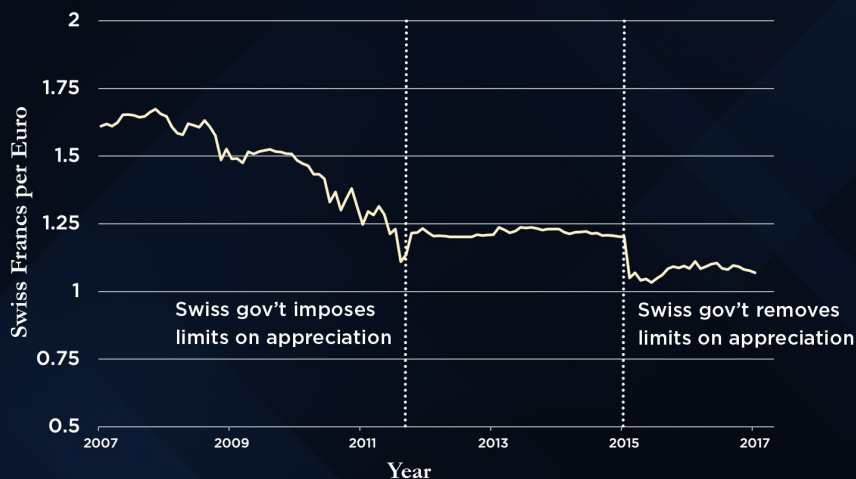
However, there are adverse longer-term implications of undervaluation. When the central bank keeps exchanging domestic currency for foreign currency, it is increasing the domestic money supply. This implies a rise in inflation as people bid up the price of domestic goods, services, and assets. For example, Japan's undervaluation of the yen after the end of the Bretton Woods system eventually led to an enormous asset bubble in Japanese stocks and land prices during the late 1980s. In addition, an undervalued currency favors exporters at the expense of importers. Both Japan in the 1980s and China in the early 2000s came under significant US political pressure with the threat of being branded as currency manipulators.

Undervaluation also means that the central bank has to manage its growing holdings of foreign exchange reserves. Since currency does not earn interest, the central bank typically has to invest this money into foreign assets. However, that can leave it vulnerable to an asset market collapse or a financial crisis in another country. For example, Singapore's sovereign wealth fund, which had invested some of the country's surpluses into the US housing market, lost 20% of its value in the early days of the 2008/2009 recession.

## 18. Why Countries Try to Control Exchange Rates

When the central bank feels pressure from negative indicators such as rising inflation or complaints from internal groups (like consumers and importers) or external groups (like unhappy foreign governments), it has to reexamine its fixed exchange rate policy. It can choose to do one of three things: First, do nothing—hold on to the existing policy and keep absorbing reserves. Second, change the fixed rate to make the currency less undervalued. This is called a revaluation. In July 2005, China revalued its currency by 2% from 8.28 yuan per dollar to 8.1 yuan per dollar and continued to revalue in small increments until it reached a value of around 6.8 yuan per dollar about three years later. Third, abandon the fixed exchange rate regime and move to a flexible regime. For example, in 2011, as the woes in the eurozone increased, money started flowing into Switzerland, resulting in an appreciation of the Swiss franc. Authorities fixed the limits of appreciation at 1.2 francs per euro. In January 2015, Switzerland announced the end of the fix. The exchange rate changed sharply by around 15% to 1.06 francs per euro and fluctuated according to market conditions after that.

The Swiss Franc-Euro Exchange Rate 2007–2017



## 18. Why Countries Try to Control Exchange Rates

How would a country choose between the three options? Consider the example of China. China saw in the 1990s that undervaluation brought economic gains in the form of more exports and higher foreign investment that overshadowed the costs of higher inflation. Thus, it chose to let the undervaluation continue. In the 2000s, when US economic pressure and unhappiness about China's exchange rate manipulation was reaching a crescendo, China tried to release some pressure with a revaluation. However, the market clearing rate was still appreciating, and reserves were still accumulating. As these pressures increased, China relented and let the currency float against the dollar.

Anticipated revaluations have their own set of problems. This is what China experienced when it revalued its currency against the dollar in 2005. Suppose you had dollars and were looking to convert them into renminbi but felt that a revaluation from the fixed exchange rate of 8.28 yuan per dollar was imminent. You would rush to convert your dollars into renminbi before the Chinese central bank revalued. Conversely, what if you had money in China and were looking to convert into dollars and you knew that the Chinese currency would become more valuable in a few days? You would wait to get your money out of China. This means that many people would be rushing to convert dollars into renminbi and that fewer people would want to convert renminbi into dollars.

The combination of a lower demand and a higher supply of dollars means that the excess supply for dollars at the fixed exchange rate will be even greater. Thus, the Chinese central bank will accumulate foreign exchange reserves and be forced to increase the domestic money supply even faster. This will make China even more uncomfortable with the current exchange rate and more likely to revalue. This will lead to an even more dramatic rise in the number of people wanting to convert dollars into renminbi and could create a self-fulfilling prophecy of revaluation. Note, however, that the central bank's hand will not be forced in the same way by having too many reserves as it will be when it is facing having too few reserves. Thus, yelling "revaluation" here leads to a stampede for the entrance.

## READING

- Bergsten, C. Fred, and Joseph E. Gagnon. *Currency Conflict and Trade Policy: A New Strategy for the United States*. Columbia University Press, 2017.
- Blustein, Paul. *The Chastening: Inside the Crisis That Rocked the Global Financial System and Humbled the IMF*. PublicAffairs, 2003.
- Conway, Ed. *The Summit*. Simon and Schuster, 2015.
- Frankel, Jeffrey. “What Three Economists Taught Us about Exchange Rate Regimes.” *Project Syndicate*, April 21, 2021. <https://www.project-syndicate.org/commentary/cooper-mundell-williamson-exchange-rate-regimes-by-jeffrey-frankel-2021-04>.
- Roberts, Richard Whitfield, William Keegan, and David Marsh. *Six Days in September: Black Wednesday, Brexit and the Making of Europe*. OMFIF Press, 2017.

# 19

## THE EURO: MANY NATIONS UNDER ONE CURRENCY?

In 1999, 11 countries joined together to form the eurozone. They gave up their own currencies and adopted a new common currency: the euro. They also created the European Central Bank (ECB) to conduct monetary policy for the entire eurozone. These countries believed that economic gains from a common currency and more reliable monetary policymaking would follow. In the two decades since the eurozone was formed, the region has grown to 19 countries. As you will learn in this lecture, despite the internal strife and design flaws that have caused much hardship in recent times for some member countries, the eurozone remains the most important development in international finance in the 21st century.



## The Exchange Rate Mechanism

After World War II, Western European nations initially moved toward joint production of coal and steel—two key components of waging war—by forming the European Coal and Steel Community. As cooperation moved beyond these two industries, the European Economic Community (EEC) was formed in 1958 by West Germany, France, Italy, Belgium, Luxembourg, and the Netherlands. Denmark, Ireland, and the United Kingdom joined in 1973. Greece joined in 1981, and Spain and Portugal joined in 1986. The EEC also expanded in scope in the early 1990s to allow for the freer movement of goods and people among the member nations, becoming the European Union (EU) in 1993.

In the early days of the EEC, during the 1950s and 1960s, many advanced economies were part of the Bretton Woods system of fixed exchange rate systems. However, even after the Bretton Woods system collapsed in the early 1970s, members of the EEC were keen

to maintain exchange rate stability with each other to support their common trade and agricultural programs. They adopted their own version of a fixed exchange rate system in 1979: the exchange rate mechanism (ERM).

In the ERM, West Germany was the leader, and the other countries were followers. Each follower fixed the value of its currency against the deutsche mark. West Germany was free to set monetary policy for itself, while the other countries had to make sure that they managed their reserves to keep their exchange rates fixed to the deutsche mark. Initially, the ERM



helped reduce inflation and provide stability to member countries. However, it collapsed in 1992 in spectacular fashion due to the reunification of East and West Germany and the ratifying of the Maastricht Treaty.

The reunification of East and West Germany became a catalyst for the collapse because the West German government invested significant amounts of money into infrastructure upgrades in the East. It also allowed East Germans to convert their effectively worthless East German marks to deutsche marks at an advantageous rate. The former East Germans were presented with both a valuable currency and a range of hitherto unavailable goods to spend that money on. Predictably, they went on a spending spree. The result was that total German consumption and government expenditure were both high. This created a burst of demand with constrained supply. East German industry could not immediately produce the types of goods that people wanted to buy, and it would take time for West German production to ramp up. Thus, inflation began to climb. The German central bank responded with a sharp increase in German interest rates, but this put pressure on other countries in the ERM. People in those other countries found it more attractive to convert their currency holdings to deutsche marks, which put severe pressure on reserves at the central banks of other member nations.

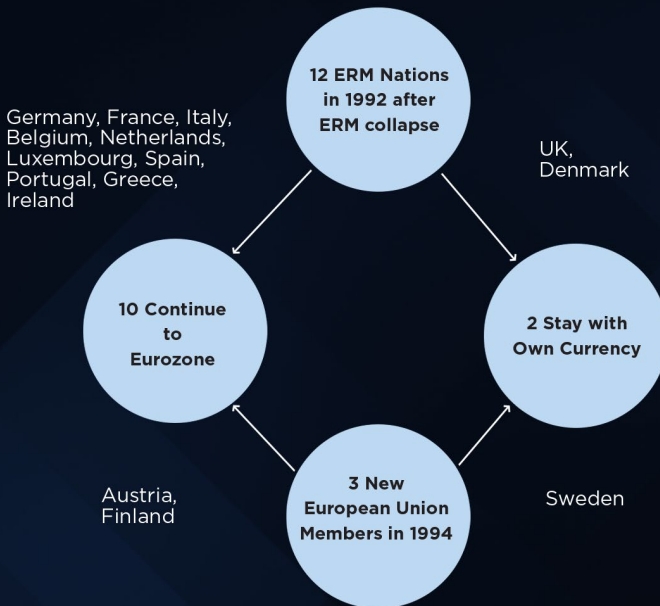
Another cause of the collapse was the fact that the Maastricht Treaty transformed the EEC into the EU and laid the groundwork toward the additional step of uniting the countries into a single currency area. The treaty was laid out in 1991 and needed to be ratified by the member countries in 1992. There were three key conditions: First, the country's currency had to remain within the ERM rates for at least two years with no devaluation. Second, the country's budget deficit couldn't be more than 3% of GDP. Third, the country's overall government debt could not exceed 60% of GDP.

Since countries were already running low on reserves in the ERM and the first condition required that countries not devalue for two years before joining the union, speculation was rampant that countries would devalue just before ratifying the treaty. Recall that once there is an expectation of devaluation because of insufficient reserves, devaluation becomes almost inevitable because of a panicked rush to the exit. Thus, in the summer of 1992, the ERM collapsed as the United Kingdom, France, Portugal, and Denmark came under attack from speculators pulling money out.

## Creation of the EU

With the dramatic collapse of the ERM, the leader-follower system came to an end. The United Kingdom and Denmark chose to maintain their own currencies. Germany, France, Italy, Spain, Portugal, Greece, Ireland, Belgium, the Netherlands, and Luxembourg decided to move to a closer union that had a new common currency that all member nations would adopt. This currency would be the euro, and the currency union would be the eurozone.

In 1994, the EU expanded to 15 countries with the addition of Sweden, Austria, and Finland. Sweden chose to stay with its own currency. Austria and Finland joined the 10 countries marching toward the euro. A date of January 1, 1999, was set for these 12 nations to meet the criteria laid out by the Maastricht Treaty and form the eurozone. By 1999, 11 of the 12 countries had done so. The member that failed to meet the criteria was Greece. It



was unable to limit government debt to be no more than 60% of GDP. Nonetheless, Greece was allowed to join the eurozone in 2001, even though its debt-GDP ratio was still higher than 100%. It qualified using a loophole from the debt clause that allowed a country to join even if its debt-GDP ratio was high, provided “the [debt-to-GDP] ratio was approaching the reference value at a satisfactory pace.”

The original 12 countries were joined by Slovenia in January 2007, Malta and Cyprus in January 2008, Slovakia in January 2009, Estonia in January 2011, Latvia in January 2014, and Lithuania in January 2015. Each country replaced its individual currency with the euro. The eurozone also created a new central bank, the ECB. The ECB was established in Frankfurt, Germany, and put in charge of forming monetary policy for the whole euro region. It focuses on euro-wide interest rates in setting monetary policy. It was also given an explicit mandate to fight inflation by keeping inflation at a rate less than 2% and given independence from political interference in achieving that mandate.

Note that the 19 member countries remain individual sovereign nations, with their own democratically elected leaders. Each nation is responsible for its own decisions about fiscal policy (both spending and taxation). Perhaps even more importantly, each country issues (and is responsible for repaying) its own euro-denominated sovereign debt. If a country has expenditures that exceed its tax revenues, it will have to issue bonds to borrow the balance. The eurozone as a whole bears no responsibility for honoring payments on those bonds; the responsibility falls on the issuing government.

## Advantages of the Eurozone

The first advantage to being in a full currency union is that a common currency removes all exchange rate unpredictability among member countries. Second is that a common currency lowers the transaction costs of moving money across the eurozone countries. Third is that by removing monetary policy autonomy from their own central banks and handing it over to the ECB, countries with irresponsible central bankers can enjoy better monetary policy outcomes—lower inflation, in particular. Fourth is that by removing depreciation risk and reducing inflation risk, the currency union lowers borrowing costs (interest rates). However, although these benefits are for the eurozone as a whole, not everyone benefits the same way.

## 19. The Euro: Many Nations under One Currency?

Roughly speaking, Germany's economy is about 30% of the entire eurozone economy, and France's comprises around 20%. Only 2 of the 19 nations make up half of the union's GDP. Italy and Spain collectively comprise another quarter of eurozone GDP, meaning 4 nations comprise 75% of the GDP. The Netherlands, Austria, Belgium, and Portugal bring this to 90% of GDP, with 8 nations. The 11 smallest countries together account for only 10% of eurozone GDP, and of those, the bottom 5 (Malta, Cyprus, Estonia, Luxembourg, and Slovenia) account for only around 1% of GDP. The differences within Europe matter in the distribution of the gains from the eurozone. Ex ante, the benefits of increased trade and investment are likely to be greater for smaller, poorer economies such as Slovenia. Having access to significant markets is already a benefit of being in the EU. The added benefit of a common currency allows economies that are poorer and have lower labor costs to attract investment from richer nations.

What about the gains from having the ECB conducting monetary policy? First, by joining the eurozone and ceding control of their monetary policy to the ECB, members are accepting the bank's explicit mandate to deliver an inflation rate of less than 2%. That is tremendously beneficial to the high-inflation economies in particular.

Second, members might get help with economic expansion. After all, to deliver on its inflation target, the ECB is also closely monitoring eurozone-wide economic outcomes. If it sees signs that the eurozone economy is slowing down, for example, it will step in with expansionary policy. However, the larger eurozone economies will tend to get more implicit attention from the ECB, simply because what happens in those economies will be more consequential for aggregate eurozone outcomes. Thus, the monetary policymaking of the eurozone is particularly beneficial for larger economies with a history of high inflation, such as Italy or Spain.

Third, joining the eurozone leads to lower borrowing costs (i.e., lower interest rates). There was substantial divergence in borrowing costs for the 11 original member nations of the eurozone plus Greece in the mid-1990s. The cost of borrowing for 10 years for Greece was close to 20% in the early 1990s compared to around 12% for Italy and around 7% for Germany. Why were the rates so different? To start, if you lent money to Greece, you would be paid back in Greek drachmas. However, if you lent money to Germany, you

would get paid back in German marks. Since Greek inflation was high, the drachma's purchasing power would fall (inflation risk), as would the value of the drachma relative to the mark (exchange rate risk). Lending to Germany was a much more attractive proposition. Moreover, Greece had much higher debt-GDP ratios than Germany. Thus, the risk that the Greek government would be unable to repay its debts was much higher (default risk).

However, as the prospects of joining the eurozone improved over time, not only did the borrowing costs fall for all countries but also the rates began to converge. Greece did not join the EEC until 1981, and its borrowing costs had not quite converged to the others' by 1999 but did so in 2001. Why? Because both exchange rate risk and inflation risk would decline. The exchange rate risk for Italian and German bonds was the same since they would all be paid back in euros. Simultaneously, the inflation risk for those bonds was the same because monetary policymaking was ceded to the ECB for both countries.

All 12 eurozone countries enjoyed the same low borrowing costs during the first half dozen years of the eurozone's existence. Note, however, that there was no transferring of fiscal policy. Each country remained responsible for spending and taxation decisions within its borders. In the prior regime, a country that borrowed too much would have to pay a much higher interest rate. For example, Greece had been paying an interest rate that was 10 percentage points higher than Germany in the early 1990s. However, now, that same Greek governmental structure was given the ability to borrow at the same rate as Germany.

Not surprisingly, Greece continued to borrow lavishly. Investors were acting as if the eurozone as a whole would become responsible for paying the debts of member countries. This misjudgment allowed countries such as Greece to borrow way more money at way cheaper rates than they would otherwise have been able to. Since these countries did not use their borrowing wisely, the access to cheap credit fueled the crisis that appeared around 2011. By the early 2010s, Greece realized that its debt burden was so great that it could no longer make payments on its loans and threatened to default. Panic ensued as creditors began to worry about other European countries that had been hit hard by the global financial crisis. Worries came to a head that the entire eurozone would unravel. Indeed, after the crisis, the monetary union did

begin to unravel. Interest rates began to diverge, partly because default risks emerged. Moreover, exchange rate risk was creeping back into the picture, following fears that Greece would leave the eurozone.

## Robert Mundell

10-Year Government Bond Yields 2007–2011



The challenges that the eurozone faced in the 2010s were anticipated 50 years prior by Robert Mundell. During the 1960s, Mundell wrote about the conditions under which a group of countries should form a currency union. One of the key criteria that he discussed was the importance of avoiding asymmetric shocks—situations where a negative shock of some type would impact only one or two countries in the currency union. The basic intuition was that when a country gives up its currency and joins a monetary union, it can no longer conduct its own monetary policy. Thus, its ability to respond to shocks is limited. This is particularly problematic when the shock is asymmetric, meaning it only affects one country as opposed to the whole region.

The conditions of the Maastricht Treaty were aimed at creating greater conformity among member countries' macroeconomic fundamentals. Those efforts were substantially successful. However, by allowing Greece to join without completely fulfilling the entry conditions, the eurozone was inviting an asymmetric shock.

Mundell also wrote about the importance of labor mobility for a currency union. The idea was that if a country is adversely affected in an asymmetric manner, the economic impact will be eased if workers can move to unaffected areas and find work. Indeed, legal work barriers have been eliminated by the creation of a single Europe. However, there remain substantial barriers to the mobility of labor due to differences in language and culture.

Mundell also listed a centralized fiscal policy as a requirement for a currency union. The intuition there was that if a country in a monetary union is hit by an asymmetric shock, then it will not be as badly off if there is a central fiscal authority. The central authority can transfer some of the tax revenue it collects from countries that are doing well to the country that is doing poorly. However, in the eurozone, there is no centralized fiscal policy. If an asymmetric negative shock hits Greece, there is no mechanism in place to use revenue from Germany or France to assist it. It also becomes a substantial political challenge to fund relief for Greece.

Note that these same arguments can be used against the United States to some degree. California and Vermont differ enormously in many dimensions, including size, type of industry, and trading partners. The United States has somewhat greater labor mobility, though that can be exaggerated. What the country has that the eurozone does not have is a strong, centralized fiscal policy. The national identity of eurozone countries coupled with the lack of mobility must mean that transferring taxes in the eurozone will pose potentially significant political problems.

## READING

Judt, Tony. *Postwar: A History of Europe since 1945*. Penguin Books, 2006.

Lane, Philip R. “The European Sovereign Debt Crisis.” *Journal of Economic Perspectives* 26, no. 3 (2012): 49–68.

———. “The Resilience of the Euro.” *Journal of Economic Perspectives* 35, no. 2 (2021): 3–22.

Marsh, David. *The Euro: The Battle for the New Global Currency*. Yale University Press, 2019.

Skaperdas, Stergios. “Myths and Self-Deceptions about the Greek Debt Crisis.” *Revue d'économie politique* 125, no. 6 (2015): 755–785.

# 20

## INTERNATIONAL TRADE AND FINANCE MUST BALANCE

In 2021, the United States ran a current account deficit of \$821 billion. Its foreign expenditures exceeded foreign income by \$821 billion, or around 4% of US GDP. Meanwhile, China ran a current account surplus that was about 2.5% of Chinese GDP, around \$300 billion. Many politicians would interpret the US current account deficit as a sign of weakness and the Chinese surplus as a sign of strength. However, the current account is only one side of the balance of payments (BOP) accounts. By the end of this lecture, you will better understand why the United States runs a large current account deficit—and why it is not necessarily bad. You will also understand what the implications are of China running a large current account surplus and why that surplus is not an unambiguously positive development either.



# The Balance of Payments

The BEA categorizes and records transactions between US resident entities and nonresident entities into three international transaction accounts within the BOP accounts. These accounts are the current account (CA), the financial account (FA), and the capital account (KA). The CA tracks activities that generate contemporaneous (current) income from abroad or require contemporaneous (current) expenditures abroad. More specifically, it records three types of transactions: trade in goods and services; income receipts and income payments on assets (e.g., stocks, loans, and bonds); and transfer payments (e.g., remittances, pensions, and foreign aid).

The FA tracks purchases and sales of assets. These assets include both foreign direct investment and foreign portfolio investment. They could include bank deposits, stocks, bonds, loans, and holdings of foreign currency. The bulk of international transactions fall into the CA or the FA. The KA basically tracks the forgiveness of existing debts. In 2020, there were about \$7 trillion worth of transactions recorded on the CA and FA as opposed to only about \$6 billion on the KA.

International transactions are entered into the BOP accounts as either a credit (+) or as a debit (-). In the CA, the accounting rule for deciding whether a transaction is a credit or debit is simple: If the transaction generates income for the United States, it is entered as a credit. If it requires expenditure from the United States, it is entered as a debit. For example, if the United States exports \$20 billion worth of beef to Russia, the transaction will be +\$20 billion to the US CA. Note that the sum of all the transactions in this part of the CA is the trade balance.

In the FA, the accounting rule for deciding whether a sale or purchase transaction is a credit or a debit is as follows: If a US entity increases its holdings of a foreign asset, then the transaction is entered as a debit. Conversely, if a foreign entity increases its holdings of a US asset, then the transaction is entered as a credit. The corollary to this rule is that sales of assets are entered with the opposite sign of purchases. For example, if the Argentinean government increases its holdings of dollar reserves by \$50 million, this will be +\$50 million into the US FA since US currency is considered to be a paper asset.

**1. Foreign direct investment**

Transaction 1: Apple computer invests \$1.7 billion into a chip manufacturing plant in Taiwan.

Current account	Financial account	Capital account
	-\$1.7 billion	

Transaction 2: Honda invests \$700 million into buying new robots for its Marysville, OH plant.

Current account	Financial account	Capital account
	+\$700 million	

**2. Portfolio investment**

Transaction 1: Warren Buffet buys \$1 billion worth of shares of Chinese cell phone company Xiaomi.

Current account	Financial account	Capital account
	-\$1 billion	

Transaction 2: A Chinese bank buys \$200 million worth of US Treasury bonds.

Current account	Financial account	Capital account
	+\$200 million	

Transaction 3: The Argentinean government increases its holdings of dollar reserves by \$50 million.

Current account	Financial account	Capital account
	+\$50 million	

Transaction 4: The Sri Lankan government borrows \$150 million from a US bank.

Current account	Financial account	Capital account
	-\$150 million	

Simply think of a loan as a transaction where the lender gives money to the borrower in exchange for a promise to repay (IOU for short). This IOU is a paper asset. For example, say the US bank has purchased a foreign asset (Sri Lankan IOU); as a foreign asset, it should be a debit to the FA.

Finally, on the less important KA, if a country receives debt relief, that is a credit to the KA. Meanwhile, a country that grants debt relief will have a debit to the KA.

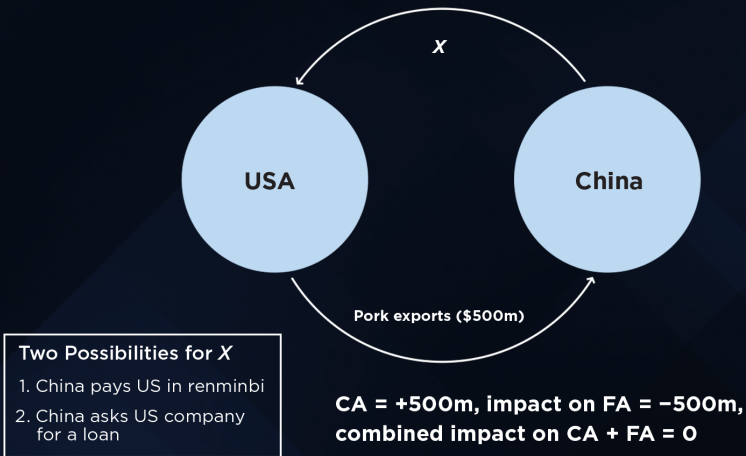
## Understanding Deficits and Surpluses

If you ignore the numerically insignificant KA and some statistical discrepancies, the following relationship must hold: CA balance + FA balance = 0. Thus, if a country runs a CA deficit (i.e., CA is less than zero), then it

must run an FA surplus (FA is greater than zero). Policymakers are often quick to praise the latter and bemoan the former. They see the CA deficit as emblematic of the failures of the US economy. However, they also see increased foreign investment into the US as a sign of strength.

What an understanding of BOP accounting teaches is that a CA deficit and an FA surplus or an FA deficit and a CA surplus are two sides of the same coin. If the United States buys more goods and services from another country than it sells to them, then that country has to be willing to accept assets from the United States in exchange; that is, every transaction from one country to another has to have a reciprocating transaction. The way in which these two reciprocating transactions get recorded ends up zeroing out in the BOP.

Suppose a Chinese company buys \$500 million worth of pork from the US producer Smithfield. This is an export and thus will enter as +\$500 million into the CA. Now, look at the other side of the transaction, labeled as  $X$ , in exchange for the pork. Here are two possible transactions for  $X$ : First, the Chinese company sends Smithfield \$500 million worth of Chinese currency; the United States has increased its holdings of a foreign asset, which is -\$500 million into the FA. Thus, impact on CA = +\$500 million, and impact on FA = -\$500 million



million. Combined impact on CA + FA = 0. Second, the Chinese company tells Smithfield that it will pay in a few months. Smithfield has effectively purchased an IOU from the Chinese company, meaning this will be  $-\$500$  million into the FA. Thus, impact on CA =  $+\$500$  million, and impact on FA =  $-\$500$  million. Combined impact on CA + FA = 0.

In 2021, when the United States ran a CA deficit of  $\$821$  billion, it also ran an FA surplus of roughly the same magnitude. In other words, the United States spent  $\$821$  billion more in expenditures on foreign goods, services, and payments abroad than it earned from selling domestic goods and services and receiving payments from abroad. Moreover, foreigners acquired  $\$821$  billion more in US assets than the United States did in foreign assets.

In essence, how large a CA deficit the United States can run is constrained only by how valued its assets are. Note that the causation can run the other way—from the FA to the CA. Suppose the South American continent is experiencing economic upheaval and uncertainty and people are moving vast sums of their savings to US banks and investing in US stock markets. This will be a surplus into the US FA. However, if the United States does not want to hold Brazilian reals, it will instead use those pieces of paper to buy something it does value—perhaps Brazilian soybeans. This will be recorded as an import and a debit into the US CA.

## Current Account Deficits

The data for the US CA from 1980 to 2021 shows that CA deficits have become the norm for the US economy since the early 1980s. The average is around 3% of GDP, with substantial fluctuations between 0% and 6%. During this period, the most severe recession was in 2008/2009. During that recession, the CA balance improved substantially, which was also true for the 2001 and 2020 recessions. In periods of strong economic recovery (e.g., the aftermath of 2001), the CA worsens. Thus, a CA deficit is not a clear signal of economic weakness.

Note that saying this is not the same as saying that a CA deficit is never a sign of economic weakness. For example, data from the World Bank used to plot Sri Lanka's CA balance shows a story that, at first glance, resembles that of the United States. Like the United States, Sri Lanka has consistently run a CA

deficit of around 2% to 3% of GDP, with the deficit occasionally fluctuating even larger. By 2020, the CA deficit was looking better than its historical average, which might mistakenly lead you to believe that the Sri Lankan economy was doing well. However, it was in deep trouble, precisely because of its international transactions.

If the Sri Lankan economy has consistently run a CA deficit, like the United States has done, then it must accordingly have run an FA surplus; that is, foreigners must have bought more Sri Lankan assets than Sri Lankans bought foreign assets. The difference lies in the details, though. Foreigners are happy to accept US dollars in exchange for goods and services. However, they do not feel the same way about Sri Lankan rupees because the rupee is not a global currency that can be used to buy commodities on global markets. Similarly, Sri Lankan companies are small and relatively unknown, meaning attracting foreign investors into the stock market is more challenging. Therefore, the Sri Lankan economy has to rely on borrowing to fund its CA deficit. Either the government or private corporations must take out loans.

The more loans they take out, the more that creditors doubt whether Sri Lanka will be able to pay what it has promised. This leads creditors to cut back on lending or raise the interest rate they charge on their loans. As the interest rates rise over time, this means that Sri Lanka's debt burden increases dramatically. This became the situation in Sri Lanka. Its CA deficit only "improved" because the country literally did not have the resources to buy imported goods, which caused tremendous economic hardship.

Thus, CA deficits can be problematic, especially for countries that do not have assets that are attractive to other countries. They are less problematic for the United States as long as the dollar remains a powerful global currency and US government bonds remain the global safe asset of choice.

## The Belt and Road Initiative

Another global macroeconomic phenomenon where the BOP sheds light is China's Belt and Road Initiative. In search of natural resources to power its manufacturing economy, China has invested vast sums of its foreign reserves in sub-Saharan Africa, Latin America, South Asia, and Eastern Europe. At first glance, this may seem like a win-win situation: China gets to transform

its excess holdings of US dollars into more productive investments in these emerging markets, giving it access to more minerals, food, or simply a higher return on investment. The receiving economies can build roads, bridges, and ports using the money they get from China.

However, whether China is helping or hurting these countries is a hotly debated topic. The infrastructure projects may solely benefit China or only enrich those in power in the recipient countries and their politically connected friends who build the road or bridge in question. In any case, the vast sums of money that the country ends up borrowing from China have to be repaid with interest.

In Sri Lanka, for example, Chinese investment is particularly noticeable in the form of new highways, a new deepwater harbor, and a gleaming new port city on land reclaimed from the sea. However, the country already could not pay back the money for the harbor and had to cede control over the harbor to China. Some worry a similar fate could befall the new port city. Simultaneously, repaying those loans adds to the CA pressures and diverts money that could have been used to buy fuel oil and cooking gas.

## Flexible Exchange Rates

Suppose that the United States has a CA balance of  $-\$400$  billion and an FA balance of  $+\$400$  billion. Now, suppose it wants to import  $\$50$  billion more in goods and services, which would move the overall BOP into deficit. This means that it has to find a way to pay for the extra  $\$50$  billion worth of foreign expenditure. Concerning the foreign exchange market, there will be an excess demand for foreign currency. US residents will suddenly be looking to acquire more foreign currency than foreign residents are looking to supply. Recall that when there is an excess demand for foreign currency in the foreign exchange market, the domestic currency (in this case, the dollar) will depreciate.

The depreciation of the dollar is what brings the BOP back into equilibrium. As the dollar becomes less valuable, this raises the relative price of foreign goods and lowers the relative price of US goods. As a result, the United States will tend to export more and import less than before. Thus, the CA balance will improve. Similarly, foreigners will be much more likely to buy US assets because of their relative cheapness, and domestic residents will be less likely to

buy foreign assets than before. As such, the FA balance will improve. This will continue until the \$50 billion deficit in the BOP is cleared and the balance ends up at zero.

Conversely, suppose that France is in BOP equilibrium with a CA balance of +€140 billion and an FA balance of -€140 billion. Then, it receives a series of foreign orders to buy €20 billion worth of French-produced submarines. This results in the CA balance improving to +€160 billion, moving the overall BOP into surplus. However, the countries that bought the submarines have to find a way to pay for the extra €20 billion. Now, there is an excess demand for euros in the foreign exchange market, meaning the euro will appreciate. This leads France to import more and export less than before. Similarly, foreigners will be much less likely to buy French assets because of their higher cost, and French residents will be more likely to buy foreign assets than before. Thus, the FA balance will worsen. The process of exchange rate appreciation will force the CA and FA balances to worsen until the €20 billion surplus in the BOP is cleared and the balance ends up at zero.

## Fixed Exchange Rates

The situation is different under a fixed exchange rate system. For instance, under a fixed exchange rate, the combination of the pandemic, political uncertainty, and the devastation from a massive explosion in Beirut led to worsening CA and FA balances in Lebanon. If Lebanon had had a flexible exchange rate, its currency would have depreciated. This would have made Lebanese goods and assets more attractive to buyers in Europe and the Middle East. Perhaps more foreign investors would have sought out investment opportunities in Beirut. However, under the dollar-based fixed exchange rate regime established at the end of 1997, excess demand for foreign currency could only result in a sharp decrease in foreign reserves at the Lebanese central bank. Lebanon's holdings of foreign reserves fell so sharply in 2020 that the central bank effectively ran out of reserves. This led to the creation of a large black market, with people exchanging Lebanese currency for dollars at rates far exceeding the official exchange rate.

## READING

- Bernanke, Ben S. “Global Imbalances: Recent Developments and Prospects.” Speech at the Bundesbank Lecture, Berlin, Germany (March 2005). September 11, 2007. <https://www.federalreserve.gov/newsevents/speech/bernanke20070911a.htm>.
- Devadas, Sharmila, and Norman Loayza. “When Is a Current Account Deficit Bad?” *World Bank Research and Policy Briefs* 130415 (2018).
- Irwin, Douglas. “Understanding Trump’s Trade War.” *Foreign Policy, Global Thinkers*. Accessed December 12, 2022. <https://foreignpolicy.com/gt-essay/understanding-trumps-trade-war-china-trans-pacific-nato/>.
- Krugman, Paul R. *A Country Is Not a Company*. Harvard Business Press, 2009.
- Schoenholtz, Kim, and Stephen Cecchetti. “Sudden Stops: A Primer on Balance-of-Payments Crises.” *VoxEU*, July 9, 2018. <https://cepr.org/voxeu/blogs-and-reviews/sudden-stops-primer-balance-payments-crises>.

# 21

## HOW STOCKS AND THE ECONOMY SHAPE EACH OTHER

**T**here is an interesting two-way relationship between financial markets and the macroeconomy. Investors in financial markets rely on having early access to good information about macroeconomic variables such as interest rates, inflation, GDP, and unemployment. The never-ending search for good information results in macroeconomic variables having large impacts on financial market outcomes. In the other direction, financial markets play a vital role in allocating capital to productive activities in the economy. Turmoil in financial markets can therefore lead to substantial negative shocks for the macroeconomy.



## Understanding Stocks

A share of stock is a claim on the ownership of the firm. Typically, the owner of a group of shares, or a shareholder in the company, owns a portion of the firm that is equivalent to her stake divided by the number of total shares issued by the company. Shareholders are therefore entitled to vote at company meetings and decide who should run the company. The extent of one's clout within the firm is proportional to the number of shares one holds. For example, Microsoft has 8.6 billion shares, of which Bill Gates owns 600 million; thus, he can cast 600 million votes.

A key source of shareholder income comes in the form of dividend payments—a portion of corporate profits that are paid out by the company to shareholders. Profits not paid out as dividends are reinvested in the company. The other source of income comes in the form of a capital gain, or a rise in the value of the shares of the firm that the shareholder owns, because other investors are willing to pay a higher price to stake their own claim to owning a piece of the company.

How does one make money on the stock market? Buy low and sell high. Consider an investor who wants to buy a share of stock, hold it for a given period of time, and then sell it. Their decision will depend on three things: the price of the stock when they pay for it now, the price they expect to be able to sell it for next period, and the dividend payments they expect to receive on the share of stock. They will be paying a price  $P_t$  and earning a dividend  $d_{t+1}$  and a sale price of  $P_{t+1}$ .

There are four key concepts to note about how to price a share of stock. First, the expectation about the dividend and stock price for the next period depends on the information you have today. Second, since the dividend and sale price are obtained next period, you need to discount those gains relative to the price you pay today. The discount rate should be higher than the nominal interest rate. A dollar in an FDIC-insured bank account next period can be discounted at the nominal interest rate. However, since any gains from the stock market are uncertain, a promised dollar from stock returns next period should be discounted more heavily than a promised dollar from your bank account.

Third, the discount rate depends on how the stock price varies with the state of the macroeconomy. Asset pricing theory says that you should discount stocks that are pro-cyclical (pay high dividends and capital gains in good times and low dividends and capital gains in bad times) more than you discount stocks that are countercyclical (pay high dividends and capital gains in bad times and low dividends and capital gains in good times). Fourth, note that stocks do not have to pay dividends at all. Many newly formed companies will often forego paying dividends since they are focused on growth. Thus, any profits they make will be reinvested into growing the business. In that case, people will be buying stocks simply for the capital gain. There is no guarantee that you can sell a stock for a higher price than you bought it for.

How do you know whether a particular share will increase or decrease in value? You need information about what the company's future prospects are. You have to buy stocks of companies whose profits will grow over time and, thus, whose shares will be desirable to many investors in the future.

## Informational Edge

In the stock market, any particular individual, bank, or investment firm could be a buyer or a seller at a point in time. Thus, good news about a company means many prospective buyers but few sellers, and bad news means many prospective sellers but few buyers. This creates the fluctuations in excess demand and supply that make stock markets volatile. Such volatility creates an insatiable demand for information. As a result, the stock market is one of the world's most amazing repositories of information and news. Every day, thousands of people look at millions of pieces of information to determine whether they should buy or sell certain stocks. To make money in a consistent fashion, they need to have access to as much information as quickly as possible.

Eugene Fama stated that the stock market is "informationally efficient" in that all relevant publicly available information will be quickly incorporated into the price of a stock. If a new piece of information about the future dividend or price of a company's stock appears, that information will be valuable to traders on the stock market. They will buy or sell the stock accordingly until the current price more accurately reflects the future dividend or stock price. All genuine profit-making opportunities are exploited, meaning only unanticipated news can move stock prices.

This search for an informational edge can have important consequences. First, economists generally believe that it is unlikely for ordinary individuals to be able to make money in a sustained fashion on the stock market because there are many other investment firms, mutual funds, and brokers who have access to more information. In “Do Investors Trade Too Much?,” Terrance Odean used data from a discount brokerage and showed that stocks that people sold tended to do better than stocks that people bought by 1.4 to 3.3 percentage points over various time horizons.

Second, the best thing for ordinary investors to do may be to entrust their money to people who have more information than they do—become a passive investor and use mutual funds or investment advisors. This creates large pools of investor funds at mutual fund companies, hedge funds, and private equity firms.

Third, the search for an informational edge can lead to illegal activity. Recall that an efficient market is one that incorporates all publicly available information. Regulators of financial markets, such as the Securities and Exchange Commission (SEC) in the United States, work to ensure that there is no insider trading on stock markets. The SEC prohibits corporate executives, directors, and other employees with access to confidential information from trading on that information or providing tips about it to others.

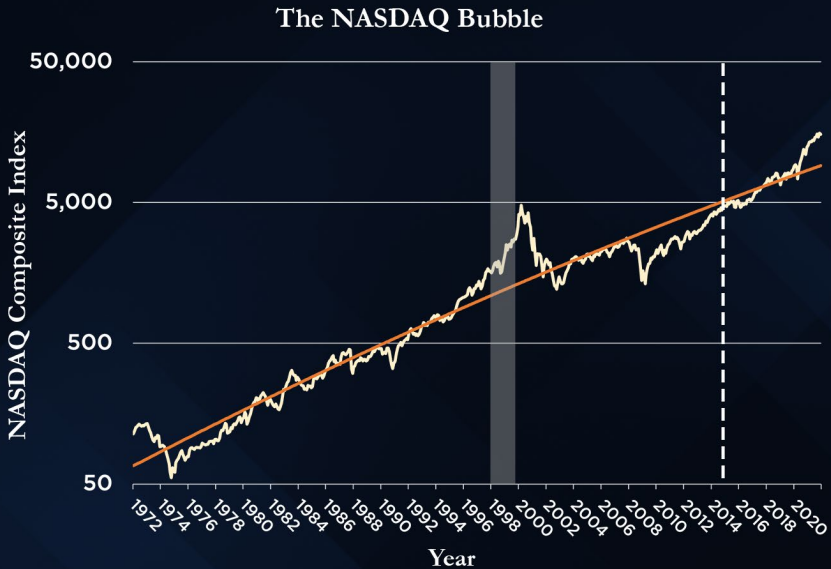
Insider information can also relate to news about the macroeconomy. Information about whether the Fed will raise interest rates or about the latest unemployment, GDP, or inflation numbers is extremely valuable and relevant for assessing the future prospects of a company. Trading on such information is also strictly restricted. A 2012 study done by Martin Scholtus, Dick van Dijk, and Bart Frijns using detailed trading data found that a delay of one second getting an order in after a (legitimate) macroeconomic good news release would result in a return that was 7.3% lower; that is, the price would have jumped higher than 7% in that first second the news came out.

Fourth, the search for information can lead to herd mentality. Individuals end up buying or selling stocks because they see others buying or selling. Stock markets see “rush to enter” and “rush to exit” scenarios all the time. As a result, they are particularly vulnerable to self-reinforcing changes in price known as stock bubbles. The term *bubble* is used to describe a movement, typically an increase, in the price of a stock that is driven solely by the expectation that someone will pay a higher price for it in the future.

## Stock Bubbles

Two of the most notorious stock bubbles of recent times were the dot-com boom in the United States in the late 1990s and the Japanese stock market bubble in the early 1990s. In the 1990s, the internet emerged as a new “information superhighway” that enhanced the ability of existing companies to reach new customers and created an opportunity for new companies to create entirely new markets. The euphoria over the endless possibilities that this new technology made available, combined with the uncertainty about which companies would be successful in this new setting, led to vast sums of money being used to purchase shares in virtually any company that promised it had a business model related to the internet in some way.

Between 1998 and 2000, the NASDAQ index, which represented the stock market valuation of tech companies, tripled in value before crashing. The danger of a bubble is that it will inevitably pop at some point. Since the rising



## 21. How Stocks and the Economy Shape Each Other

price is not based on any fundamentals (actual profits that the company is making), the price rise is entirely predicated on new buyers emerging to keep driving the price higher.

Suppose you were a lucky early buyer of shares in an internet company at \$2 a share and the price has since risen to \$20. Now, you have increased your stake tenfold. However, the person who bought at \$20 will need the share price to go to \$200 to have the same gain you did, which may be considerably harder to do. At some point, the share price will plateau as new buyers go in search of other, more attractively priced stocks. When the price plateaus, early investors will now decide to cash out. When they sell their shares, the stock price starts to fall. Then, herd mentality can lead to even more sales, and the price falls faster. When the NASDAQ bubble popped, within a year, it had plummeted to around a quarter of the peak value. It took almost 14 years before it returned to the peak that it reached in 2000.

The Japanese stock market bubble was even more astounding. In the 1980s, overexuberance regarding the prospects of Japanese companies led to sharp increases in Japanese stock market prices. This was reflected in the Nikkei

### The Japanese Stock Bubble



index increasing fourfold in the space of four years in the late 1980s. The Japanese stock market—and the real estate market, which had seen a similar unsustainable increase—collapsed suddenly in the early 1990s. A quarter century later, the Nikkei was only half of what it had been at the peak of the bubble. After 30 years, the Nikkei index still has not returned to its prior heights.

From a macroeconomic perspective, the suddenness with which the bubble collapses and the severity of the collapse have a devastating effect on the economy. The stock market collapse in Japan led to a massive recession and almost three decades of a near-stagnant economy. The United States experienced a similar event in the Great Crash of 1929 that precipitated the Great Depression.

# How Stock Prices Affect the Economy

The primary channel through which macroeconomists study the impact of the stock market on the economy is through wealth effects—how changes in the value of stock holdings impact a household's wealth and influence its consumption. This is difficult to empirically estimate because investors may cut back on stock purchases when they hear news about bad economic conditions that may be forthcoming. This leads to a situation where a fall in stock prices precedes a recession but does not necessarily cause the recession.

Using data from the period 1990 to 2015, Alp Simsek and coauthors have estimated that a dollar of lost stock market wealth results in a fall of around 3 cents of consumption. Thus, in 2008/2009, when the value of US equities declined by around \$12 trillion, that reduced consumption spending by almost \$400 billion, around 3% to 4% of GDP. Karl Case, working with Robert Shiller and John Quigley, showed that wealth effects are even greater for housing markets: 5 to 9 cents of consumption change per dollar change in housing wealth. For most middle- to lower-income Americans, their home is the only substantial asset they have. This explains why the 2008/2009 crisis was so devastating, since both stock market and housing market wealth were destroyed.

When the economy is booming, stock markets and housing markets rise in value, and consumer wealth increases. However, if it's their wealth that is rising rather than their income, consumers are likely to be financing their consumption by borrowing more. For example, if your wealth rises from \$350,000 to \$620,000, you don't get your hands on that \$270,000 unless you sell your stocks or your house. Most people probably continue to hold on to the asset rather than selling it. The additional consumption comes from borrowing against that increased wealth. When the housing or stock market collapses, reducing wealth, consumption will fall because of the wealth effect and because the money borrowed in good times has to be paid off.

Another channel through which stock prices can affect the economy is investment. The most famous theory about what drives investment in the macroeconomy is known as Tobin's  $q$  theory. James Tobin defined  $q$  as a ratio of market value to book value. The market value of a firm is the value of all its outstanding shares, and the book value of the firm reflects what could be obtained if all of the firm's assets were sold off at market prices. If  $q$  is large for a company, then that company has incentive to invest even more because every dollar of additional investment raises its stock price by  $q$  dollars. Tesla, for example, had a market value worth \$1 trillion at the end of March 2022 compared with a book value estimated at around \$25 billion. This gave it a Tobin's  $q$  of around 40. This enabled Tesla to invest more in larger factories, expand into more markets, and buy up competitors using its valuable shares.

This effect is compounded by the balance sheet channel of lending. When a firm approaches a lender, the lender is concerned about adverse selection. This refers to the lender's self-doubt. However, concerns about adverse selection will tend to dissipate with a high stock market valuation. After all, if the stock market is the hub of information aggregation in the economy, then a high market valuation must reflect good news about the future of the company and, hence, make the lender more willing to lend.

Both Tobin's  $q$  and the balance sheet channel magnify economic fluctuations. In good times, stock market valuations rise, and companies are more likely to have a high Tobin's  $q$  and expand their investment. They are also more easily able to borrow money from lenders. Thus, economic activity booms. In

contrast, in bad times, stock market valuations fall, and companies are likely to have a lower Tobin's  $q$ . They invest less and find it difficult to borrow, and economic activity falls.

Policymakers understand the links between the stock market and variables such as economic consumption and try to behave in a predictable way to avoid spooking the stock market. When there is an unexpected downturn in the economy, the Fed usually steps in and takes action. It not only undertakes the standard expansionary monetary policy by lowering interest rates but also takes steps to make sure that financial markets keep functioning properly. The additional liquidity that the Fed provides to banks helps stem the decline in the stock market, as investors do not have to sell their shares to come up with cash to meet their needs. Thus, part and parcel of good macroeconomic policy is to act predictably and communicate well.

### READING

Case, Karl E., Edward L. Glaeser, and Jonathan A. Parker. "Real Estate and the Macroeconomy." *Brookings Papers on Economic Activity* 2000, no. 2 (2000): 119–162.

Kindleberger, Charles. *Manias, Panics, and Crashes: A History of Financial Crises*. Palgrave Macmillan, 2015.

Malkiel, Burton. *A Random Walk down Wall Street*. W. W. Norton, 2020.

Shiller, Robert J. *Irrational Exuberance*. Crown Business, 2006.

Wood, Christopher. *The Bubble Economy: Japan's Extraordinary Speculative Boom of the '80s and the Dramatic Bust of the '90s*. Solstice Publishing, 2005.

# 22

## LABOR MARKET POLICIES FOR THE LONG TERM

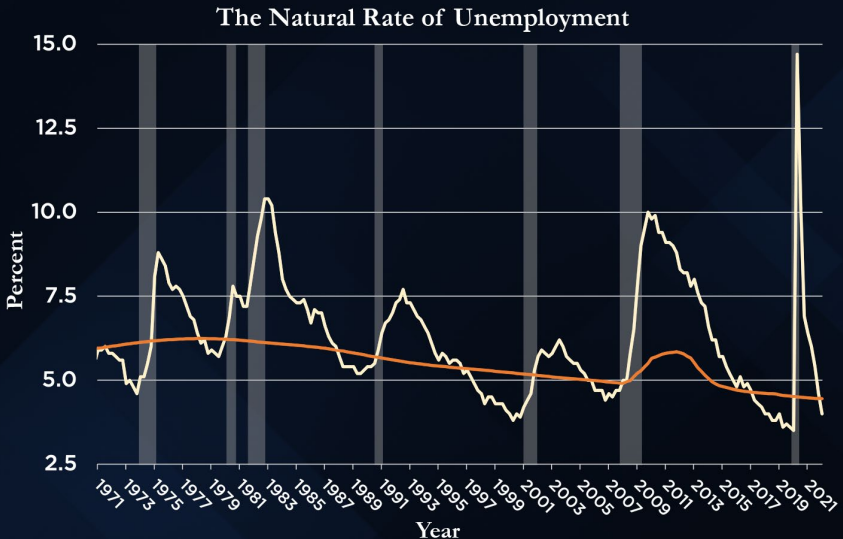
**W**hen would it be helpful (or possibly harmful) for the minimum wage to change? How can policymakers decide when (and how much) to increase unemployment benefits to mitigate inequalities in a recession? What should be done when technological progress and globalization in the modern economy seem to be leaving some workers in certain industries behind? This lecture will illustrate that a deeper understanding of how the state of the macroeconomy impacts the labor market helps address such policy questions.



# The Natural Rate of Unemployment

Milton Friedman introduced the concept of a natural rate of unemployment to refer to the unemployment rate that exists in the absence of short-run fluctuations. The term *natural rate of unemployment* simply means that there's a longer-term baseline level of unemployment determined by the structure and condition of labor markets. Note that this rate cannot be directly measured. Instead, it is inferred from the data on unemployment rates or derived from a model of unemployment. There is considerable variation in what its precise estimate is.

One calculation of the natural rate that economists pay close attention to is an estimate done by the CBO. The actual unemployment rate fluctuates around the natural rate. Moreover, the natural rate is not a fixed number. The CBO revised its natural rate steadily upward during the 1970s and then steadily downward for the next three decades to below 5%. Strikingly, the disruption of the 2008/2009 recession lasted long enough that the CBO raised its estimate of the natural rate of unemployment.



## 22. Labor Market Policies for the Long Term

The difference between the unemployment rate and the natural rate is the cyclical unemployment rate. This is related to the short-term cyclical fluctuations in the economy. There's an inverse relationship between the state of the economy and the cyclical unemployment rate. When the economy is booming and unemployment is low, it will be less than zero. Conversely, when the economy is in a recession, it will be positive. How significant is this effect? For the United States, Okun's Law says that for every 2 percentage points that GDP rises above (or falls below) its long-term growth trend, the cyclical unemployment rate will fall (or rise) by around 1 percentage point. The knowledge macroeconomists have about policies that can reduce the frequency and magnitude of economic fluctuations will therefore reduce (by half as much) the spikes in unemployment rates caused by cyclical unemployment.

In a typical recession, GDP will fall by about 4 percentage points relative to its long-term trend. Okun's Law predicts that in such cases, the cyclical component of unemployment will be 2 percentage points (i.e., the unemployment rate will be 2 percentage points higher than its long-term trend). Even in 1981/1982, when GDP fell by around 8 percentage points, this relationship seemed to hold: Cyclical unemployment rose by around 4.6 percentage points.

During the pandemic, however, it became clear that Okun's Law is not an ironclad law of nature. In the second quarter of 2020, GDP fell by about 11% relative to its growth trend. However, cyclical unemployment rose by more than Okun's Law would have predicted, to around 8%. This was a much higher impact than in a typical recession because of the nature of the public health threat posed by the pandemic. There was both a labor demand shock and a labor supply shock.

However, cyclical fluctuations are only one part of the story. Between January and April of 2020, the unemployment rate went from 3.5% to 14.8% overall, but this varied with education. For high school dropouts, it went from 5.7% to 21%, whereas for people with a college degree, it increased from 2% to 8.4%. Over a longer horizon, this is the normal state of affairs for these two groups in the labor market.

## How Unemployment Rates Vary

The baseline rate of unemployment declines with age. Overall, 16- to 19-year-olds consistently have a much higher natural rate of unemployment than 20- to 24-year-olds. The latter in turn consistently have a higher rate than those who are 25 to 54 years old. The lowest rate of all belongs to those older than 55.

Baseline unemployment rates also vary to a lesser degree by gender. In the 1970s, women seemed to have a higher natural rate of unemployment, but over time, the gap disappeared. In fact, after the 2008/2009 recession, male unemployment rates were higher than female unemployment rates. Concerning unemployment rates by race or ethnicity, African American and Hispanic unemployment rates tend to be considerably higher than white and Asian unemployment rates.

Improving labor market outcomes for all groups requires the identification of policies that can improve the natural rate of unemployment. This requires a deeper understanding of that rate and its main drivers. Economists describe the existence of a natural rate of unemployment as stemming from frictional unemployment and structural unemployment.

## Frictional and Structural Unemployment

Frictional unemployment occurs because it takes time to make connections between unemployed workers who have skills and experience to offer and employers who have vacancies that require those skills and experience. Structural unemployment, however, occurs when the skills and experience of unemployed workers do not match up with the needs of the vacancies that firms are trying to fill. Often, this is because of technological advances or changes in the structure of the economy.

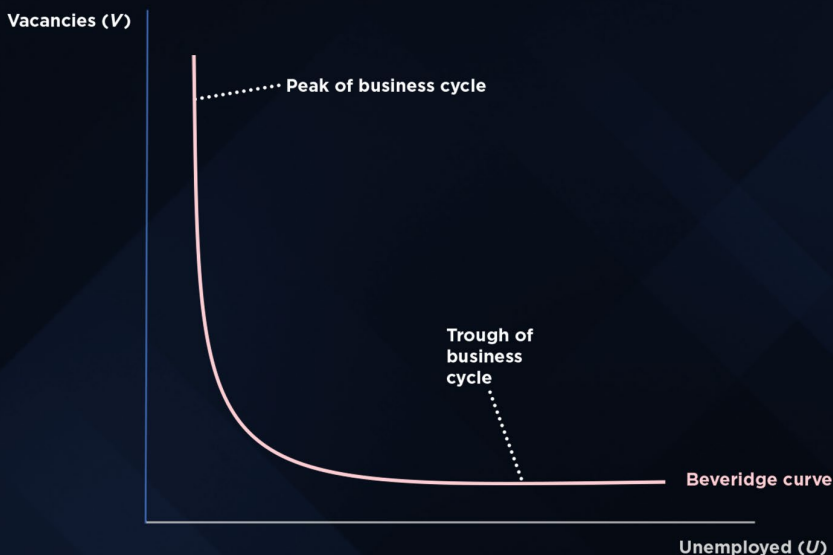
The key difference between frictional and structural unemployment is that frictional unemployment is short-lived. It simply takes time for the workers to find the employers whose needs match their skills. However, for a structurally

## 22. Labor Market Policies for the Long Term

unemployed worker to find new employment, there has to be a more substantive change in circumstances—either a reversal in economic fortunes that reboots to bring back old industries or, more likely, a retraining or re-skilling of the workers to make them more able to find work in a different occupation or industry.

Peter Diamond, Dale Mortensen, and Christopher Pissarides built labor market models that highlight the importance of incorporating job search and job matching in finding the equilibrium of the labor market. Their models draw on an empirical relationship between the number of unemployed workers and the number of vacancies that is known as the Beveridge curve, stemming from the work of William Beveridge in the 1940s.

In theory, the Beveridge curve should be a downward-sloping, concave relationship between the number of vacancies and the number of unemployed workers that is consistent with a given natural rate of unemployment. When economic activity is at the peak of a business cycle, it is expected that there are fewer unemployed workers and many unfilled jobs because firms are



struggling to find qualified workers who want to work at prevailing wages. However, at the trough of a business cycle, firms post few vacancies, and a large number of unemployed workers are trying to find work. Thus, you move down along the Beveridge curve during the transition from peak to trough as a recession occurs, and back up along the Beveridge curve during the transition from trough to peak as the recovery takes hold.

The relationship is nonlinear: At the depths of a recession, the number of vacancies is already low. Thus, unemployment can increase rapidly because matches are hard to make. Moreover, at the peak of a business cycle, when there are few unemployed workers, the number of vacancies can rise rapidly because matches are hard to make. Changes in the rate at which these connections between unemployed workers and vacancies are made are reflected in shifts of the Beveridge curve.

Constructing an empirical Beveridge curve was challenging during the 20th century because accurate data on the overall number of vacancies in the economy was not easily available. This led the BLS in December 2000 to institute a new survey of firms called the Job Openings and Labor Turnover Survey. More than 15,000 firms are surveyed monthly and asked about new hires, loss of existing workers, and unfilled positions. This new survey defines a vacancy as an unfilled position that is seeking someone to work in the next 30 days, with the firm actively recruiting from outside of its own establishment.

## Better Information, Training, and Education

Enhanced access to information on vacancies and applicants can help speed up the hiring process on both sides of the labor market. This will shift the Beveridge curve inward, reducing frictional unemployment and lowering the natural rate. Governments can play a role by setting up job-finding centers that can help unemployed workers link up with firms that have job vacancies. However, access to more information does not have to necessarily be government-funded. Greater use of the internet through sites such as LinkedIn not only helps the unemployed find jobs faster but also helps people search for a new job even before they leave their current position.

Moreover, increased financial resources for retraining and re-skilling programs available to support workers in industries adversely impacted by trade or technology changes are vital in reducing structural unemployment. David Autor and his coauthors have shown that geographic areas with a heavy concentration of jobs in sectors that were exposed to greater competition from Chinese imports suffered heavy job losses after US trade with China expanded in the 1990s. These labor market shocks remained even a decade later. This indicated that the workers who lost their jobs due to Chinese competition were unable to find jobs in other sectors that may have boomed as a result of enhanced trade with China.

Although the US government funds a program called Trade Adjustment Assistance to help fund training and job searching for workers displaced by trade, other work by David Autor has shown that the program does not help workers successfully reenter the labor market. Many of these displaced workers end up relying on Social Security disability benefits to eke out a subsistence level of living. The China shock is an example of a negative shock that would shift the Beveridge curve outward. Improved retraining and assistance programs for workers displaced by trade and technology can shift the Beveridge curve to the left.

There is also a substantial gap in labor market prospects of highly educated workers versus less-educated workers. Accordingly, substantial labor market benefits can result from policies that increase the quantity of education at all ages and stages of life. However, in the United States, local funding of secondary education means that trade and technology shocks that hit particular sectors and communities hard also have damaging effects on the education (and future job and income prospects) of the children of laid-off workers. Finding a way to break this link could help improve labor market outcomes in the long run and move the Beveridge curve to the left.

Such support for education might include reducing high school dropout rates and providing vocational or trade school opportunities for those not wanting to continue their education in college. It might also include increasing support for the quality of education in regions with downturns, such as by emphasizing STEM education to take advantage of skill-biased technological change.

## Reducing Payroll Taxes, Unemployment Benefits, and Labor Market Regulations

In the United States, employees and their employers each have to pay half of a 12.4% Social Security tax and a 2.9% Medicare tax, for a total payroll tax rate of 15.3%. Reducing or eliminating the employer portion of the payroll tax could encourage more firms to hire more workers. Furthermore, reducing or eliminating the employee portion would encourage more people to seek work since they would get to keep a larger slice of their wages. Higher payroll taxes can lead to more use of informal employment, in which people work jobs that are off the books. Workers in these positions may have less bargaining power. Some of these jobs are likely to be in illegal activities, and some are in occupations where needed health and safety precautions are more likely to be violated.

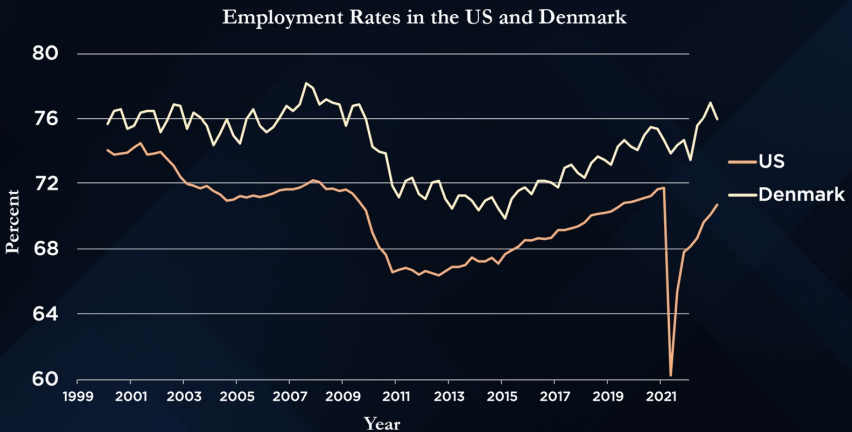
However, in the United States, payroll taxes are used to fund Social Security and Medicare payments for the elderly. If these taxes are cut without other offsetting changes to the budget, the government will either have to raise income or corporate taxes or cut back in other areas of spending, like education or health. These changes could offset some of the positives that come from lower payroll taxes.

Another policy is to reduce excessive unemployment benefits. The pandemic provided a good example of the double-edged sword that is unemployment benefits. In the initial stages of the pandemic, the federal government increased the amount of benefits, lengthened their duration, and expanded the scope of who was eligible to receive them. By cushioning the impact of a recession on household budgets, allowing the newly unemployed time to look for a new job, and helping people deal with health care and childcare costs, unemployment benefits may help stop frictional unemployment from turning into structural unemployment. When the government extended these programs in 2021, however, employers and some governors felt that the generosity of the enhanced benefits was leading workers to turn down job offers to return to work.

## 22. Labor Market Policies for the Long Term

Identifying what constitutes an excessive unemployment benefit is tricky. Economists have studied what happened later in the pandemic, when nearly two dozen states cut off the expanded unemployment benefits for their workers. They compared what happened to workers in these states to what happened to workers in states that did not cut off benefits. They found that there was a 4% increase in employment—but also a 35% decrease in how many people received benefits. This suggests that most people removed from benefits did not find employment.

The design of the unemployment benefits program also matters. In the United States, unemployment benefits are designed to deal with the fallout from frictional unemployment, with a standard duration of 26 weeks. In contrast, in Denmark, unemployment benefits have a two-year maximum period and are coupled with aggressive investment in retraining programs to address the longer-term challenges of structural unemployment. Denmark's overall percentage of the working-age population employed during the 21st century is consistently higher than that in the United States.



Labor market regulations can also move the Beveridge curve. Regulations such as health and safety standards, sick leave, and overtime pay for work beyond a certain number of hours in a week can greatly enhance the quality of life for workers. That can encourage more workers to enter lines of work that they may otherwise deem to be too risky. Providing workers more flexibility in choice of work and greater confidence regarding their wages can reduce the likelihood of workers leaving the job market. However, with too many regulations, firms may be less likely to create vacancies and much more cautious about filling them. Frictional unemployment may increase if firms face more constraints, such as on their ability to hire or fire workers.

### Lower Minimum Wages

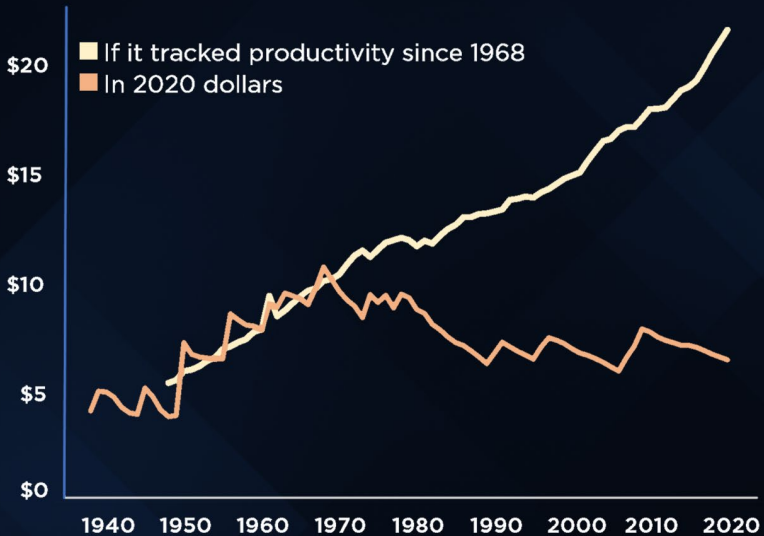
The standard theoretical argument is that a binding minimum wage can increase unemployment because during times of hardship, a firm may not have to let workers go if it can pay them a bit less. For example, in the early days of the pandemic, suppose a restaurant was located in a city where the minimum wage was \$15 per hour. Even if the restaurant could theoretically keep all its staff at \$10 per hour, it would not be able to do so because the minimum wage prevents it from paying \$10 per hour. Higher minimum wages can also be a barrier for less-experienced workers to find jobs because firms may not want to take on untested workers at high wages. Thus, higher minimum wages can reduce the efficiency of the matching process and shift the Beveridge curve out.

Some economists argue that higher minimum wages need not increase the natural rate of unemployment and could even end up reducing it. For example, a higher minimum wage that enables a worker to take care of basic needs can make them more likely to keep working in the face of childcare needs. Firms may also pay more attention to hiring higher-quality workers and invest in training and retaining workers if they have to pay a higher wage. This would shift the Beveridge curve in.

Economists who have looked at state and local minimum wage increases have found that such increases in recent times have not had a detrimental impact on employment. The most authoritative recent study, done by Arindrajit Dube and covering the period from 1979 to 2016, estimated that the federal minimum wage could be increased by around \$3 from current

## Trends in the Minimum Wage

Minimum wage, in 2020 dollars and if it had kept pace with productivity, 1938–2020



levels without causing a noticeable increase in unemployment. This is because the real minimum wage is low in relation to both historical wage rates and productivity rates. The United States first established a minimum wage in 1938, and it kept up with inflation and productivity for 30 years, through 1968. However, the inflation-adjusted minimum wage has since declined from \$12 per hour in 1968 to roughly \$7 per hour by 2022. Moreover, it has fallen far below the rate of productivity gains in the economy, which, by one estimate, might correspond to a minimum wage of \$24 per hour.

## READING

Autor, David, David Dorn, and Gordon Hanson. “The China Shock: Studying the Impact of China’s Rise on Workers, Firms and Markets.” Accessed December 12, 2022. <https://chinashock.info>.

Brauer, David. “The Natural Rate of Unemployment.” Working Paper 2007-06, Congressional Budget Office, 2007. <https://www.cbo.gov/publication/18566>.

Dube, Arindrajit. “Designing Thoughtful Minimum Wage Policy at the State and Local Levels.” The Hamilton Project, June 16, 2014. [https://www.hamiltonproject.org/papers/designing\\_minimum\\_wage\\_policy\\_at\\_state\\_and\\_local\\_levels](https://www.hamiltonproject.org/papers/designing_minimum_wage_policy_at_state_and_local_levels).

Munté, Júlia Rivera. “Policy Brief: Spain: Youth, COVID-19, and the Struggle of Unemployment.” European Student Think Tank, February 16, 2022. <https://esthinktank.com/2022/02/16/policy-brief-spain-youth-covid-19-and-the-struggle-of-unemployment/>.

Petrongolo, Barbara. “The Diamond, Mortensen and Pissarides Nobel.” VoxEU, October 15, 2020. <https://voxeu.org/article/diamond-mortensen-and-pissarides-nobel-search-and-market-frictions>.

# 23

## RISING INEQUALITY IS A THREAT TO ALL OF US

In the United States, the COVID-19 pandemic widened the gap between the rich and poor. A 2019 Census Bureau report showed that US income inequality was already at its highest point in 50 years. Although the US economy recovered quickly, sharp rises in the cost of fuel and food implied even greater pressures on those with low incomes. Simultaneously, some of the richest people in the world were able to vastly increase their wealth. This lecture will take a closer look at how economists measure inequality and discuss some explanations for rising inequality that researchers have advanced.



## Understanding Income Inequality

Economists agree that it's important to consider poverty, mobility, and opportunity when studying economic inequality. Richard Wilkinson has shown that the psychological effects of inequality take a heavy toll in the form of stress, illness, substance abuse, and subsequent increases in crime and incarceration. In the United States, income inequality among the adult population almost certainly contributes toward income inequality for children.

Regarding opportunity, cutting-edge economic research by the Harvard economists Raj Chetty and Nathaniel Hendren through their Moving to Opportunity project has shown that when a family is given a housing voucher that allows them to move to a better neighborhood, the outcomes for young children in eventual adulthood are much better than those for similar-aged kids in families left behind. In other words, the bad adult outcomes for kids in poor neighborhoods reflect inequality in opportunity.

Recall that data about people's income in the United States is collected by the US Census Bureau in a monthly survey of about 60,000 households called the CPS. The key variable in the CPS is the income of a household, which comprises all related family members and unrelated individuals who live in the same housing unit.

The most commonly used data is money income. This is pretax earnings from wages, pensions, Social Security payments, unemployment benefits, interest, rental and dividend income, cash welfare programs, and educational benefits. Household income here does not include capital gains or in-kind benefits that are conditional on low income. The omission of the increase in assets is necessitated by the fact that people do not usually incur the capital gain until they sell the asset. Thus, the income of high-earning households is likely to be significantly understated by the CPS.

## Calculating Income Inequality

Households can be ordered from poorest to richest, and the resulting distribution can be summarized by dividing the households into equal-sized groups. The custom is to divide into 10 groups, called deciles, each containing

## Income Distribution of Households in the United States

Quintile	2020 HH Income Limit	Decile	2020 HH Income
		Lowest 10%	15,600
Lowest 20%	27,026	Lowest 20%	27,026
		Lowest 30%	39,535
Lowest 40%	52,179	Lowest 40%	52,179
<b>Median income</b>		Lowest 50%	67,521
Lowest 60%	85,076	Lowest 60%	85,076
		Lowest 70%	107,908
Lowest 80%	141,110	Lowest 80%	141,110
		Lowest 90%	201,126
Lowest 95%	273,739	Lowest 95%	273,739

**Mean income: \$97,000**

10% of households; or into 5 groups, called quintiles, each containing 20% of households. In 2020, there were approximately 130 million households in the United States. The 13 million households with the lowest income earned below \$15,600, while the 13 million households with the highest income earned above \$201,126. The median household income was \$67,521, while the mean income level was \$97,000. When the mean is higher than the median, this means income dispersion is higher in the upper ends of the income distribution. Note that the CPS does not report income beyond the 95th percentile: Because it is a survey, there is no guarantee that it can accurately capture the handful of households at the top of the income distribution.

Three quick and easy measures of inequality are the 90/10 ratio, the 90/50 ratio, and the 50/10 ratio. The 90/10 ratio shows how the 90th percentile and above of households do compared to the bottom 10%. That is \$201,126 divided by \$15,600, or about 12.9. So the richest 10% of households earn about 13 times as much as the poorest 10% of households do.

### 23. Rising Inequality Is a Threat to All of Us

Remember that \$15,600 was the upper limit for the bottom 10%, while \$201,126 was the lower limit for the top 10%. The 50/10 ratio shows how the bottom 10% of households compare to the median household:  $\$67,521 \div \$15,600 = 4.33$ . The median household earns at least 4.3 times what the poorest 10% of households earn. Similarly, the 90/50 ratio shows how the top 10% of households compare to the median household:  $\$201,126 \div \$67,521 = 2.97$ . The richest 10% of households earn about 3 times what the median household earns.

These ratios can be tracked over time to see how the relative distribution has changed. The gap between the top and bottom households has been rising steadily over time. The gap between the top and median households has also been increasing, but less dramatically. The gap between the median household and the lowest 10% declined from 1970 to 1980 but has stayed fairly constant over the past 50 years. Overall, this suggests that rising inequality in the United States is driven more by the rich doing better than the typical household over time than by the poor doing worse than the typical household over time.

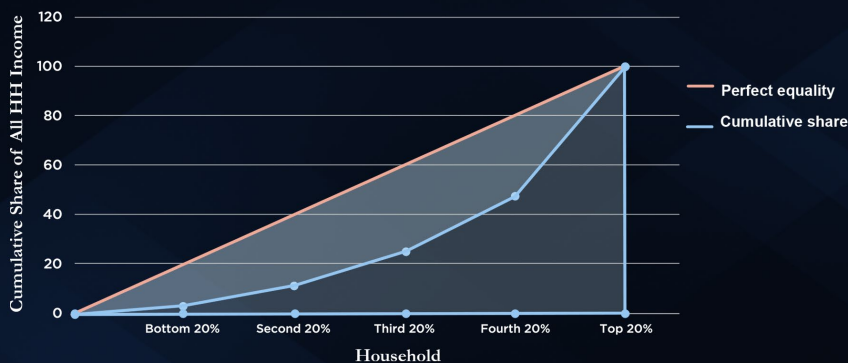
Another way to measure income inequality is to compare the income of all households in a particular decile or quintile to the total income of all households in the country. In 2020, according to the CPS, the bottom 20% of households by income earned only 3% of total income for the United States. In contrast, the top 20% of households earned more than 52% of total income. The top 5% alone earned about a quarter of all household income.

## The Lorenz Curve and Gini Coefficient

The above data is typically presented in a relationship called the Lorenz curve. In a Lorenz curve, households are ordered by income level on the horizontal axis. The cumulative share of all household income earned by households at or below that income level is on the vertical axis. In addition to the actual distribution of income, the Lorenz curve diagram also includes a 45-degree line that shows what the income distribution would look like under perfect equality (or zero inequality): The bottom 20% would earn 20% of household income, for example. The area between the 45-degree line and the Lorenz curve is a measure of inequality. The more unequal the distribution of income, the larger the area.

## 23. Rising Inequality Is a Threat to All of Us

Lorenz Curve for the United States 2020



The Gini coefficient is the most widely used measure of income inequality in economics. It ranges in value from zero to one. A value of zero indicates perfect equality—a theoretical extreme in which every household earns exactly the same. A value of one, however, reflects the other theoretical extreme: All of the income is earned by the richest household in the country. A higher Gini coefficient indicates greater income inequality. The Gini coefficient for the United States has increased steadily over the past 50 years, indicating that income inequality has been increasing.

Lorenz curves and Gini coefficients can be calculated for different countries or groups of countries. Data from the Organisation for Economic Co-operation and Development shows that the United States has a higher Gini coefficient than most developed nations. Scandinavian countries have much more equality in income than the United States. However, some economies have even higher levels of inequality. South Africa, with its legacy of apartheid, has one of the most unequal distributions of income in the world.

When looking at the changing patterns of income distribution over time, it is important to keep certain factors in mind. First, note that the CPS surveys do not track the same people as years pass. For example, considering income for the lowest 20% of households in 2015 and 2020, some of the households in the lowest 20% in 2015 may have moved to a higher grouping in 2020.

Economic research on the United States shows that about two-thirds of the people in any one quintile move to another quintile over a 10-year period. About half of those in the top quintile move to a lower quintile, and about half of those in the bottom quintile move to a higher quintile.

Second, changes in household composition can impact household income. If two individuals who were living separately form a household, then household income doubles. However, if one adult leaves the family, the income of the household declines. About half of the increase in the rising Gini coefficient in the United States is due to rising divorce rates together with marriage taking place at later ages.

Third, there is a difference between the pretax and the posttax distribution of income. The United States has a progressive income tax system, where the marginal income tax rate on wage income rises as income increases. Furthermore, it has means-tested transfer programs, or programs that provide additional income to low-earning workers that get phased out as their income increases. Thus, economic inequality after taxes is likely to be less, particularly when comparing the poor to the median household. However, taxes on capital income tend to be lower than taxes on wage income, meaning posttax inequality between the rich and the middle class will likely be higher posttax than pretax.

## Why Is Income Inequality Rising?

David Autor, Lawrence Katz, and Melissa Kearney did one of the most comprehensive studies of wage inequality in the United States from the 1960s to the early 2000s. They looked at what they called the polarization of economic inequality—whereby there is a rise in inequality in the top half of the distribution and a more stable level of inequality in the bottom half. Their conclusion was that education differences play a significant role. Specifically, the earnings gap between workers with a postgraduate degree and those who lack a college degree widened much more rapidly than did the gap between college-only graduates and those without a high school degree.

Another explanation is changes in the composition of labor resulting from skill-biased technological change. During the 1990s, employment in the occupations with the highest and lowest levels of education expanded, while

### 23. Rising Inequality Is a Threat to All of Us

employment in the middle-educated occupations shrank. This reflects what economists sometimes refer to as the hollowing out of the American middle class as a result of skill-biased technological change. Middle-income workers were replaced by a combination of automation, computers, and skilled workers. However, low-income service jobs occupied by less-educated workers were much less affected.

A third explanation, particularly relevant from the 1990s onward, is globalization—or, more specifically, the rising prominence of China. David Autor has shown that areas in the United States that were more exposed to competition from Chinese imports saw a dramatic contraction in employment that was persistent. Since competition from China was primarily in manufacturing, the types of jobs that were lost were jobs closer to the middle of the income distribution.

Emmanuel Saez and Gabriel Zucman have done exhaustive data work to show that the rise in income inequality at the top overwhelms any trends seen in the more visible parts of the income distribution. The average growth rate of income per adult from 1946 to 1980 was around 2% a year. This 2% growth was fairly evenly experienced across the income distribution. In fact, only the richest individuals did not enjoy this growth: Their growth rates were closer to 1%. In contrast, from the post-1980 period, most incomes grew at 1.4% per year. Within this slower growth rate, things were much more unequal. The bottom 85% of the income distribution experienced growth slower than 1.4%. In fact, the bottom 10% experienced negative growth. In contrast, the growth of income increased sharply for the top 1% and at a fantastically sharp rate for the top 0.0001%.

The economic inequality within the top 1% mirrors the economic inequality in the population at large; that is, the gap within the top 1%, or between its internal top 10% and bottom 10%, has grown at least as much as the gap between the top 10% and the bottom 10% of the overall population. The same would be true for those within the top 0.1% of the distribution: The 99.99% to 100% of the distribution will be doing better than those between 99.90% and 99.91%.

# Superstar Economies and Supermanagers

What can explain the rise of this phenomenon since 1980? Sherwin Rosen coined the term *superstar economy* to capture the forces at work. The argument is as follows: Globalization has raised the size of the market for certain professions. For instance, entertainment is watched by billions of people worldwide, but they all want to see the best of the best. Only a handful of actors can command \$25 to \$50 million a movie. Top athletes can earn tens or hundreds of millions of dollars a year. Factoring in outside endorsements, the gap between newer talent and prior talent becomes even more dramatic.

An alternative explanation for the rising incomes at the top comes from Thomas Piketty. In *Capital in the Twenty-First Century*, Piketty argues that the largest change is in the small number of high-level executives in large firms, or supermanagers. The crucial change was a move toward compensating these supermanagers with stocks and stock options as opposed to cash. Changes in corporate governance structures have allowed these executives to set high salaries for themselves and also set the salary structure in such a way that the rewards have no downside risk. For example, sales at the department store Kohl's dropped 20% during the COVID-19 pandemic, yet the CEO, Michelle Gass, still earned \$13 million—almost a 40% increase from the previous year.

More surprisingly, Piketty also argued that compensation of the supermanagers is driven by luck more than skill. Consider the compensation for top hedge fund managers, who pool together investors' funds and use those funds to make investments. The typical compensation structure for these funds is called two and twenty: The manager gets to keep 2% of the funds collected by the investors plus 20% of the gains on those funds. This compensation structure is designed to give outsized gains with little downside risk. In years where the stock market does well, the hedge fund manager collects 20% of gains in addition to the 2% of the funds. In years where the market does poorly, the manager still collects 2% of the funds without having to share in the losses. Overall, the rewards to supermanagers, especially in financial occupations, have exceeded the rewards to top figures in other occupations.

### 23. Rising Inequality Is a Threat to All of Us

Solutions to income inequality are often discussed but rarely implemented. The broad-based transfer programs of the COVID-19 pandemic briefly reduced economic inequality. However, they were quickly demonized as being wasteful, discouraging work, and driving inflation, and they were not renewed in subsequent budgets. Inequities in public education are acknowledged on all sides, but neither political party has been able to convince the other that their preferred policy should be implemented on a large scale. Attempts to raise tax rates on capital or on high-income workers or inheritances or wealth run into a gauntlet of political opposition, funded by the richest people and corporations with the most to lose. As inequality increases, the distrust of the status quo also increases, creating political upheaval. The trends of the past 40 years should not continue for the next 40 years. If they do, the United States will approach the extremes of inequality currently found in countries such as Mexico or South Africa.

## READING

Milanović, Branko. *Global Inequality: A New Approach for the Age of Globalization*. Belknap Press, 2019.

NPR. “Moving to Opportunity?” August 30, 2019. <https://www.npr.org/transcripts/756028025>.

Piketty, Thomas. *The Economics of Inequality*. Belknap Press, 2015.

Rodrik, Dani. “A Primer on Trade and Inequality.” Working Paper 29507, National Bureau of Economic Research, November 2021. <https://www.nber.org/papers/w29507>.

Saez, Emmanuel, and Gabriel Zucman. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. W. W. Norton, 2019.

Yglesias, Matthew. “Everything You Need to Know about Income Inequality.” *Vox*, May 12, 2015. <https://www.vox.com/2014/5/7/18076944/income-inequality>.

# 24

## THREE MACROECONOMIC CHALLENGES OF OUR TIME

**T**his course concludes by focusing on three universal economic challenges for which macroeconomists need to devise solutions: income inequality, migration, and climate change. These issues are at the frontier of macroeconomics, and sometimes it is not clear how even macroeconomists who have spent decades studying the economy will agree on definitive answers to them.



## Income Inequality

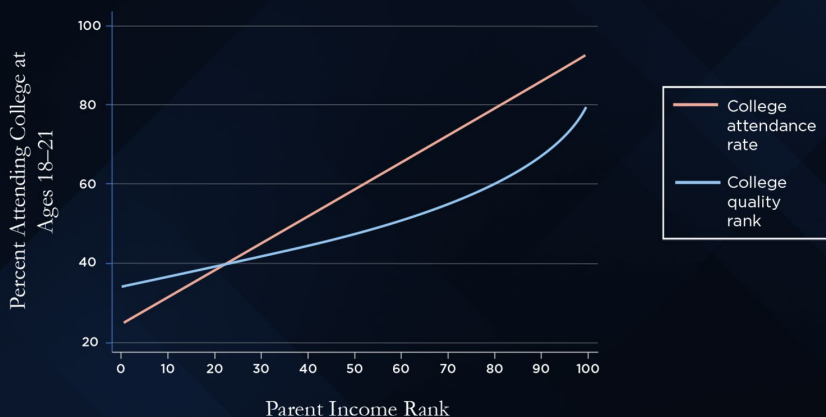
The first challenge for macroeconomists concerns how to create more equitable economic growth. Miles Corak came up with a compelling relationship known as the Great Gatsby Curve. It compares each country's Gini coefficient (a measure of economic inequality) with a measure of intergenerational mobility (how unrelated an individual's earnings are to their parents' earnings). A higher value on the horizontal axis (the Gini) shows more economic inequality in a country, and a higher value on the vertical axis shows more income mobility across generations.

There are a few things to note here. First, there is a downward-sloping relationship: Greater inequality is associated with reduced mobility. Second, some of the most unequal economies in the world are the more prominent emerging economies (e.g., India, South Africa, and Chile). Third, the highest mobility seems to be in Scandinavian nations rather than in the United States. Increasingly, the evidence has shown that the United States lags behind many other rich nations in this area.

How can the economy become less unequal and more mobile without sacrificing its high level of per capita GDP and sustained economic growth? The simple answer is that everyone needs to benefit from technological progress. That requires better education for the current generation and unlinking the educational outcomes for the next generation from the relative prosperity of the current generation. Raj Chetty has shown how incredibly strong the correlations are between college attendance and family income as well as between college quality and family income. He highlights that it is people at the upper end of the income distribution who typically attend elite higher education institutions.

If those with higher educational attainment are those who benefit the most from growth but achieving higher educational attainment is linked to parental income levels, then more equitable growth in the future will require that society break this tight persistence of educational inequality. However, achieving more equitable growth also means better safety nets and greater redistribution to help those who cannot take advantage of technology-driven economic growth. During the pandemic, income transfers to the unemployed

## Children's College Attendance Rate and Quality vs. Parent Income Rank



and low-income parents were passed by Congress. These programs were estimated to have cut child poverty by 30%, shored up consumer spending, and helped make this the shortest recession in recent history.

## How Taxes Could Help Reduce Inequality

Most economists hope for agreement on two areas related to taxes that can both reduce inequality and increase mobility. The first is the idea of an income tax with the following properties: a broad tax base with low average tax rates, marginal tax rates that rise with income, and marginal tax rates that apply to capital and labor income the same way. The current system has some of these features but falls behind on others. For example, 12.4% Social Security taxes are assessed on the first \$150,000 of income but not on earnings beyond that. This lowers the tax base and raises the average tax burden on those who make less than \$150,000. Multinational corporations

can move corporate income and profits all over the world to avoid paying taxes in the United States, but small businesses located in the United States can't. This reduces the tax base, and the smaller companies pay a 21% tax rate that is higher than what larger multinationals pay.

There are progressive tax rates on wage income, but capital income is taxed less than labor income. This results in situations where there is a marginal tax rate of 22% on a teacher who makes \$75,000 and a higher rate of 37% on an anesthesiologist who makes \$700,000 a year, but then a marginal tax rate of only 15% on a hedge fund manager who makes \$7 million a year because most of their income is considered to be capital income.

Second is the idea that large bequests should not be allowed to be passed on to future generations without being taxed. The United States currently does not tax the first \$12 million of a taxpayer's estate. This means that the estate tax is not relevant for the entire bottom 99.9% of the US population. Above that limit, the current tax code provides many loopholes that allow for passing on intergenerational wealth.

One change in particular that economists have offered is to get rid of the step-up in basis of an asset at the time it is transferred to the next generation. Suppose you purchased \$10,000 worth of shares in Apple in 1994; that is considered the cost basis for those assets. If you sold these shares in 2022, they would be worth \$1 million. You would have to pay a capital gains tax of 15% on the \$990,000 difference between your \$1 million sale and your \$10,000 basis. You would pay around \$150,000 in taxes. However, if you died and passed on your wealth to your children, then the starting basis for those shares would be recalibrated to be \$1 million (i.e., the value at the time they inherited it). If your children sold the shares for \$1 million, they would owe no taxes.

One group of macroeconomists has been closely studying a different idea related to taxation—adding a tax on wealth as a means of reducing existing income inequality. Emmanuel Saez and Gabriel Zucman calculated in 2019 that the top 0.1% of the US income distribution holds around \$12 trillion in wealth. Thus, a 1% marginal tax on wealth would raise about \$120 billion, which is about 3% of existing tax revenues. Adding a wealth tax would greatly increase the progressivity of the US tax system. However, economists broadly

disagree on whether such a tax would be practical to implement. Wealth held in the form of art, mineral rights, intellectual property, or privately held corporations is difficult to value.

Economists have also looked at the possibility of creating a universal basic income (UBI). Ideas for a UBI have typically proposed an unconditional guaranteed fiscal transfer payment from the national government to every citizen. For budgetary reasons, these proposals typically would replace all existing safety net programs with the UBI. However, when Hilary Hoynes and Jesse Rothstein studied UBI proposals for developed countries, they concluded that the proposed programs would end up providing more support to middle-class households than to poor households. Moreover, they would provide less support to households with children, elderly family members, and disabled family members than existing programs do.

## Migration

The issue of inequality also interacts with migration. Rich countries with aging populations and low population growth are seeking new workers to keep the labor force growing and fill in gaps in sectors where there is a shortfall of domestic workers. In addition, their aging societies have placed strains on safety nets that require younger, healthier workers to pay taxes into the pension systems that protect the old. However, rich countries worldwide are seeking ways to block the arrival of people from poor countries searching for jobs and opportunity.

Michael Clemens, in the *Journal of Economic Perspectives*, estimated that if 5% of the people living in the poorest countries in the world were able to move to wealthier nations, the global economic gains would exceed the benefits of removing all barriers to trade and capital flows combined. In *Let Their People Come*, Lant Pritchett calculated that if rich countries allowed enough poor-country migrants in to increase their labor force by 3%, this would lead to \$300 billion in higher income for the migrants and \$51 billion in increased income to the residents of the rich countries. Instead, he shows, rich nations would rather spend \$17 billion on measures to limit migration and \$70 billion in aid to alleviate poverty in the nations that the migrants would come from.

## 24. Three Macroeconomic Challenges of Our Time

What might an economically sensible migration policy look like? Economists who study this issue have outlined four key features of a feasible system. First, there has to be a focus on the countries that are the most likely source of migrants. Thus, a policy that focuses on high-skilled immigration alone would not be adequate for the United States. Immigration reform must address the single largest source of pressure: poor workers from Mexico and Central America.

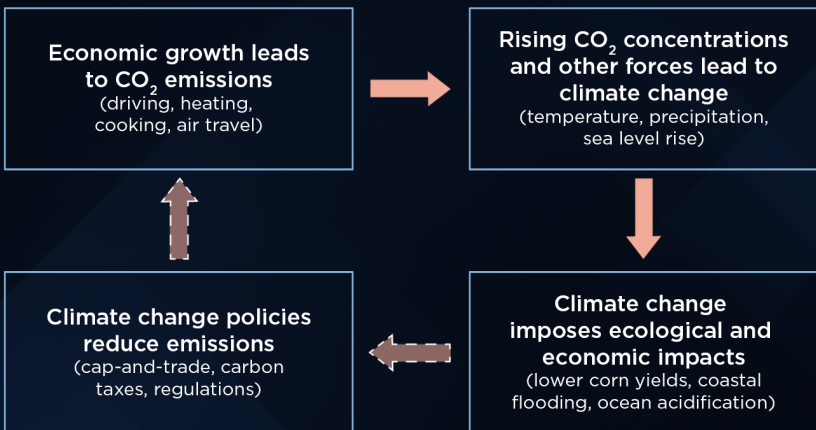
Second, there needs to be a points-based system that prioritizes occupations, skills, or industries that the country needs more workers in. This is in contrast to the system in the United States in which the immediate relatives of an individual who is either a US citizen or has legally migrated to the United States are prioritized for immigration purposes. According to data from the Department of Homeland Security, in the five years before the pandemic hit, of the 5.5 million immigrant visas granted, only around 700,000 (around 13%) were based on employment-related criteria.

Third, temporary work permissions with no path to citizenship or permanent residency can be more successful in that they are more politically palatable. In the United States, one such program is the H-2A visa program that allows employers to bring in agricultural workers for seasonal work. An individual who gets a work permit under this program has to renew the permit every year and agree to spend three months outside of the United States every three years. The permits are only issued for workers from approved countries—that is, countries that have cooperated on immigration-related matters.

The views of macroeconomists on the migration issue are most reflected by the Immigration Modernization Act of 2013. This act provided more worker permits for both low-skilled and high-skilled workers; priority for international students who receive a STEM degree from a US academic institution; pathways to citizenship for people who were already in the United States but without proper legal status; and resources to implement a verification system for employers to ensure that employees had legal clearance to work. However, the bill did not pass the House.

# Climate Change

Changes in global climate will have a staggering impact on global income inequality. The effects will fall heavily on some of the poorest people in the poorest countries in the world, such as the 175 million people in the low-lying land of Bangladesh. In his Nobel Prize lecture from 2018, William Nordhaus discussed a major challenge ahead: Economic growth will lead to increased emissions. The higher emissions will result in greater carbon dioxide concentrations, causing climate change. Climate change imposes costs on economies, which leads the impacted countries to implement policies to combat climate change. This in turn will impact economic growth. Nordhaus pointed out that in the absence of concrete emissions reduction agreements or policies, it is unknown how the steps to reduce emissions will impact the scale and nature of economic growth.



People in rich countries live lives that emit more greenhouse gases and contribute more to global climate change than people in poor countries. The World Bank estimates that there are 1.2 billion people living in high-income nations, with per capita GDP exceeding \$12,500. Meanwhile, there are 4 billion people living in lower-income and lower-middle-income economies

with per capita GDP under \$5,000. If sustained economic growth transforms the lower-income populations into higher-income populations, then the pressures on climate change will increase.

Moreover, climate change has implications for economic fluctuations as well as for economic growth. In late 2020, the Fed began taking steps to assess the risks posed by climate change to the stability of the financial system. In a 2021 speech, Fed vice chair Lael Brainard discussed two specific threats: First is the increased likelihood of natural disasters that result in widespread damage to property and threaten both the ability of insurers to honor claims and the ability of uninsured property owners to bear losses. Second is that the collateral for many real estate loans held by banks and financial institutions will be at risk from increased sea level rise and flooding.

Carbon emissions that contribute to climate change are a classic example of a negative externality. The public costs to society from the emissions of a factory far exceed the private cost to the factory of producing those emissions. In other words, emissions are a “public bad” rather than a public good—a non-rival, non-excludable cost on everyone. The classic economic remedy here is to raise the market price of emissions to match their social cost. This can be done via a carbon tax imposed on either emitters, who must pay for each ton of greenhouse gases they emit, or energy users, based on their consumption. Alternatively, the market price could be increased with a system of tradable permits. The rights to emit greenhouse gases would have to be purchased from a limited supply of such rights. In the absence of carbon taxes, the price of carbon emissions is zero.

Nordhaus provided estimates that taxes that start at \$45 per ton of carbon emissions would be an optimal rate to balance the costs of additional carbon emissions with the benefits of the resulting production. However, insofar as the cumulative effect of past emissions practices must be undone, more must be done to reduce emissions; thus, the price of carbon emissions would have to be raised. Nordhaus estimated that if the goal is to stop global temperatures from rising by more than 1.5 degrees in 100 years, the price would be \$260 per ton. In terms of consumer prices, a \$45-per-ton carbon tax would imply around a 61-cent-per-gallon tax on gasoline. A \$260-per-ton carbon tax would be around a \$3.60-per-gallon tax on gasoline. Higher carbon taxes will have economic and political repercussions in the short run: They will serve as a negative supply shock to the economy, much like any other run-up in gasoline prices.

However, recognize that imposing costs for carbon emissions does not necessarily mean sacrificing long-run economic growth. Pricing carbon so that the cost of burning it matches the social cost it imposes could spur technological innovation to reduce carbon usage. Similarly, technological progress could help mitigate and overcome the repercussions of past emissions. Furthermore, a higher carbon tax could generate revenue that could be used to offset other taxes, fund research into technological advancements to help manage climate change, and compensate regions and communities adversely impacted by climate change.

In the end, there is a strong macroeconomic argument that there should be no tension between the need for sustained economic growth and lower greenhouse gas emissions. Both require the same thing in the long run: technological progress. As Paul Romer showed, technological progress requires getting the incentives right. In the case of scientific research, incentives came in the form of intellectual property rights. In the case of carbon emissions, incentives can come in the form of getting the price of carbon emissions right.

### READING

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. “Race and Economic Opportunity in the United States: An Intergenerational Perspective.” *Quarterly Journal of Economics* 135, no. 2 (May 2020) 711–783. <https://opportunityinsights.org/paper/race/>.

Kahn, Matthew. *Adapting to Climate Change: Markets and the Management of an Uncertain Future*. Yale University Press, 2021.

Kurtzleben, Danielle. “How Would a Wealth Tax Work?” National Public Radio, December 5, 2019. <https://www.npr.org/2019/12/05/782135614/how-would-a-wealth-tax-work>.

Nordhaus, William. “Climate Change: The Ultimate Challenge for Economics.” Nobel Prize Lecture, December 8, 2018. <https://www.nobelprize.org/uploads/2018/10/nordhaus-lecture.pdf>.

Pritchett, Lant. *Let Their People Come: Breaking the Gridlock on Global Labor Mobility*. Brookings Institution Press, 2006.







## **Copyright © The Teaching Company, 2023**

IMAGE CREDITS: Getty Images (pages 9, 43, 91, 158),  
Brandmeister/Wikimedia Commons/Public Domain (page 177)

Printed in the United States of America

This book is in copyright. All rights reserved.

Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of The Teaching Company.

4840 Westfields Boulevard, Suite 400

Chantilly, VA 20151-2299

USA

1-800-832-2412

[www.thegreatcourses.com](http://www.thegreatcourses.com)